A Publication of the American Psychological Association

# The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

Articles

Is the F Scale Irreversible? ..... 143
*Richard Christie, Jean Havel, and Bernard Seidenberg*

Differential Autonomic Responsiveness in the Process-Reactive Classification of Schizophrenia ..... 160
*Gerald F. King*

Ambiguity and Anxiety ..... 165
*Andrew S. Dibner*

The Application of Learning Theory to the Treatment of Tics ..... 175
*Aubrey J. Yates*

CSR Conditioning in Anxiety States, Normals, and Chronic Functional Schizophrenic Subjects ..... 183
*Edmund S. Howe*

Parental Evaluations of Children and Children's Self-Evaluations ..... 190
*Malcolm M. Helper*

Identification in Group Therapy ..... 195
*Gerald J. Briskin*

Some Relationships Among Projective Test, Interview, and Sociometric Measures of Dependent Behavior ..... 199
*Bernard J. Fitzgerald*

*Continued on inside front cover*

Volume 56 MARCH 1958 Number 2

CONTENTS—*Continued*

The Repression Hypothesis Studied in a Situation of Hypnotically Induced Conflict ..... 204
*Ruth A. Bobbitt*

The Effect of Sibling Competition and Alliance on Level of Aspiration, Expectation, and Performance ..... 213
*Eva Dreikurs Ferguson*

Effects of Public and Private Failure on Self-Evaluation ..... 223
*Ezra Stotland and Alvin Zander*

Distance as a Determinant of Influence in a Two-Person Serial Interaction Situation ..... 230
*Seymour Fisher and Ardie Lubin*

The Influence of Language-Acquisition Contexts on Bilingualism ..... 239
*W. E. Lambert, J. Havelka, and C. Crosby*

Obligations to Self and Society in the United States and Germany ..... 245
*David C. McClelland, J. F. Sturr, R. H. Knapp, and H. W. Wendt*

Sex Differences in Problem-Solving Performance as a Function of Attitude Differences ..... 256
*Gloria L. Carey*

Personality and IQ Change ..... 261
*Jerome Kagan, Lester W. Sontag, Charles T. Baker, and Virginia L. Nelson*

The Psychology of Smoking ..... 267
*Charles McArthur, Ellen Waldron, and John Dickinson*

CRITIQUE AND NOTES

Attitude Change and Justification for Compliance ..... 276
*Arthur R. Cohen, Jack W. Brehm, and William H. Fleming*

Anality, Political Aggression, Acquiescence, and Questionnaire Construction ..... 278
*Maurice L. Farber*

Independence and Resistance to Perceptual Field Forces ..... 279
*Douglas N. Jackson*

Extrapunitiveness and Ethnic Attitude ..... 281
*Gerald S. Lesser*

Anticipated Audiences as Determinants of Recall ..... 282
*Wilbur Schramm and Wayne Danielson*

ERRATA ..... 284

Volume 56 MARCH, 1958 Number 2

# The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
*Harvard University*

URIE BRONFENBRENNER
*Cornell University*

ROGER BROWN
*Massachusetts Institute of Technology*

DONALD T. CAMPBELL
*Northwestern University*

LAUNOR F. CARTER
*System Development Corporation*

RICHARD CHRISTIE
*Columbia University*

MORTON DEUTSCH
*Bell Telephone Laboratories*

I. E. FARBER
*State University of Iowa*

LEON FESTINGER
*Stanford University*

DONALD W. FISKE
*University of Chicago*

ROBERT R. HOLT
*New York University*

IRVING L. JANIS
*Yale University*

HAROLD E. JONES
*University of California*

HAROLD H. KELLEY
*University of Minnesota*

LEONARD S. KOGAN
*Community Service Society of New York*

DANIEL J. LEVINSON
*Harvard Medical School*

GARDNER LINDZEY
*University of Minnesota*

DAVID C. MCCLELLAND
*Harvard University*

LEO POSTMAN
*University of California*

ELIOT H. RODNICK
*Duke University*

NEVITT SANFORD
*University of California*

SEYMOUR SARASON
*Yale University*

RICHARD L. SOLOMON
*Harvard University*

RENATO TAGIURI
*Harvard University*

Managing Editor: ARTHUR C. HOFFMAN

Circulation Manager: HELEN ORR
Editorial Assistant: BARBARA B. CUMMINGS

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Graduate Department of Psychology
New York University
Washington Square
New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1958 by the American Psychological Association, Inc.

# IS THE F SCALE IRREVERSIBLE?[1]

RICHARD CHRISTIE
*Columbia University*

JOAN HAVEL
*Cornell Medical School*

AND BERNARD SEIDENBERG
*New York University*

ONE of the most frequently used instruments in the area of personality and attitude measurement is the F scale developed by the authors of *The Authoritarian Personality* (1). It has not always been used judiciously, but many data have been collected in conjunction with it in hundreds of samples. It has been found to be correlated (significantly) with attitude measures varying from Arab nationalism (24) to Xenophobia (6) (not to neglect Xenophilia! [22]) and with other variables as diverse as Necker cube reversals (15) and amount of out-group aggression in dreams (19).

The F scale was designed as a covert measure at the personality level of incipient fascistic tendencies, although it is more commonly interpreted as a measure of authoritarianism.

A feature of the F scale that has come under recent critical scrutiny is the unilateral wording of its items. These are so phrased that agreement with them is scored as indicative of potential fascism, disagreement as nonfascistic. Most of these criticisms stem from Cronbach's summary of earlier work on response set and his indication of its importance in answering test or scale items (10, 11). Some subjects, it appears, have general tendencies to accept or reject items. Although this susceptibility to respond in a set way may be indicative of a broader factor of acquiescence, it need not manifest itself uniformly. According to Cronbach, response set is more likely to occur when a given item is vague or otherwise lacking in specificity or is in an area with which the individual is unfamiliar. Such considerations are pertinent to many of the items in the F scale which are characterized by nonspecificity of referent and do not make sense when analyzed logically.

## RESPONSE SET AND THE F SCALE: SOME QUESTIONS

Although the authors of *The Authoritarian Personality* gave more detail about their research procedures than is customary, they failed to anticipate all the critical questions that might be raised and there are explanatory lacunae. One of these is the failure to explain why F-scale items were unilaterally worded. This is in contrast to an earlier discussion of the explicit rationale for a similar wording of the E (Ethnocentrism)-scale items since it was recognized that this violated recommendations of experts on scale construction (1, p. 59).

Three of the 38 items in the original Form 78 F scale were worded in such a fashion that rejection would imply fascistic tendencies (Nos. 12, 20, and 28: 1, Table 1 (VII), pp. 226–227). Two of them discriminated negatively (Nos. 12 and 20) and ranked 37th and 38th in rank discriminatory power in the item analysis (1, Table 3 (VII), p. 245). These two items were believed to be poor because they were "unclear or ambiguous" (1, p. 247) and were dropped from further revisions of the F scale. The third reversal (No. 28) had poor discriminatory power and was not included in subsequent revisions of the F scale. It does emerge worded in the *opposite* direction, i.e., in conformity with the rest of the F-scale items, in Form 60 A of the F scale.

The difficulties in constructing satisfactory

[1] This study would not have been possible without the generous aid of those cited below. Donald T. Campbell, N. L. Gage, and Douglas N. Jackson made prepublication copies of their articles available and kindly furnished further details of their research. Tentative reversals of F-scale items were critically scrutinized by Robert Holt, Marie Jahoda, Robert Lee, and Gudman Smith and their advice was in large part accepted. Angus Campbell, Richard I. Evans, Warren Miller, Lester W. Milbrath, A. Robert Rollin, and Peggy Cook played a crucial role in recruiting subjects and their aid cannot be fully acknowledged in a footnote. Finally, the services of Stanley Budnitzky, Daniel Clastner, Ethelreda Seidenberg, and Orlando Sepulveda in scoring questionnaires and performing computations are gratefully acknowledged.

item reversals of the F scale that are hinted at in this initial attempt are highlighted by four recent studies in which F-scale items were reversed. Correlations between the scales composed of reversed items and the original F scale (or an abridgment) indicate little success in constructing adequate reversed items. If scales composed of these reversed items measured "authoritarianism" defined as the rejection of democratically worded items it would be expected that the correlation between them and the original F scale, which was based upon agreement with authoritarian or nondemocratic statements, would approach +1.00—if subjects are ideologically consistent and the reversals are adequate. This is clearly not the case in these four studies. Bass found a correlation of but +.20 (2), Chapman and Campbell ones of +.17 and −.01 in two samples (7), Jackson, Messick, and Solley one of −.35 (14), and Leavitt, Hax, and Roche ones varying from −.66 to +.42 in five small samples (17).[2]

Four independent studies leading to findings of negligible if not negative correlations between F scale and scales composed of reversals of its items must be taken seriously. The lack of high positive correlations was attributed to the influence of response set by all four groups of investigators. Bass even went so far as to assign three-fourths of the "reliable" variance and 59 per cent of the total variance to response set (2, p. 622). This conclusion has been challenged by Messick and Jackson who pointed out that Bass misidentified his factors when they were reflected and that a factor analysis was not justified given the nature of Bass' data (20). Although Bass acknowledged the validity of this criticism he still argued that acquiescent response set is an important component in results obtained with the F scale (4), as did his two critics.

Such an evaluation is at least partially consistent with other data. Cohn (6) was one of the first to consider seriously the possibility that the F scale was in part a measure of acquiescent tendencies. He found a correlation of +.41 between a "plus" scale of diverse items from the MMPI and an unspecified version of the F scale. Bass mentions finding only positive correlations among 13 scales of 20 aphorisms each although the content areas covered by the scales varied (2, p. 617). Items from these scales appear to be used in a scale of "social acquiescence" which is reported to correlate +.49, +.48, and +.16 respectively with the F scale in three samples (3, p. 297). Chapman and Campbell report obtaining a "rather pure" measure of response set by scoring 74 questionnaire items that were balanced in positively and negatively worded direction for a general tendency to agree. This measure was correlated with the F scale to the extent of +.43 (7, p. 131). Leavitt, Hax, and Roche cite an unpublished study by Mullery in which correlations between the F scale and acceptance of contradictory maxims of +.46, +.41, and +.38 respectively were found in three small samples (17, p. 215). A subsequent study by Carey, Rogow, and Farrell (5) on the relationship between agreement with positive and negative aphorisms and the F scale indicated that in three of four samples, agreement with both sets of aphorisms was significantly correlated with F-scale scores (positive aphorisms were defined as affirming Christian, humanist, or democratic values; negative aphorisms were those judged to deny these ideals). Webster found that a measure of response set correlated +.26 with the F scale (26, p. 31) among Vassar students.

These findings are congruent with those of Newcomb (21) in the Bennington study. He had heard Thorndike speak of the "tendency to agree" and although he was not familiar with any studies analyzing what Cronbach was later to call response set he analyzed the P.E.P. (political-economic progressivism) scale in terms of this "tendency to agree" and found that the split-half reliability on the P.E.P. scale among seniors in the class of 1939 was +.68 when scored by subtracting the number of disagreements from the number of agreements (21, p. 194). The correlation between tendency to agree as Freshmen and Seniors was +.33 (21, p. 195). These and other findings reported by Newcomb are of particular interest in the present context since the P.E.P. scale was a modification of Stagner's pioneering scale of fascist attitudes (21, p. 20).

[2] The sign of the correlations in these studies has been standardized in this discussion so that a positive correlation indicates consistency of content and a negative one a lack of relationship between the F scale and a reversal scale scored so that rejection of the items is scored as authoritarian.

Thorndike had actually conducted research on the tendency to agree during the 1930's. This material has not been published but an abstract of a paper given by Lorge on the topic at a regional psychological convention exists (18). Cronbach also summarizes additional material on reliabilities of scales in these studies (11). Briefly, split-half reliabilities for given response categories were high (.88 and .95), and an unspecified positive correlation was found between the number of "yes" responses on the Bernreuter and the number of statements checked on a Thurstone scale.

The preceding studies are similar in that the measure of response set is based upon statements that may vary in the extent to which they are ambiguous to respondents. In other words, the presence of response set is confounded to a certain degree with the item content.

Gage, Leavitt, and Stone have used a different technique to measure response set. They used true-false information items in a test pitched at such a level of difficulty that their respondents could not achieve better than a chance score. The split-half reliability of an acquiescent set, i.e., the number of "true" responses, was +.68. (This is identical with that found by Newcomb on the P.E.P. scale and suggests hypotheses about the nature of the political and economic attitudes of his Bennington sample.) A similar scale constructed by Gage *et al.* had items at a 75 per cent level of difficulty and the reliability was but +.09, a result indicating that "... *difficult and ambiguous items are required to elicit the acquiescent set.* ..." (12, p. 98, italics added).

Gage *et al.* also administered, among other measures, Cohn's "plus" scale and the F scale. The latter correlated +.44, a finding similar to Cohn's. The correlation between their Information-True (items at chance level) test and the F scale was +.35. This strikes us as the most clearcut indication of the relationship between the F scale and a response set measure that is free from variance attributable to the attitudinal content of the items. Roughly an eighth of the variance in response to the F scale would then be attributable to response set—at least among studies conducted on the usual samples of American college students (the lower correlation obtained by Webster is believed to be attributable to the fact that his sample was probably more intellectually sophisticated than members of the other samples reviewed).

Although the weight of evidence indicates that responses to the F scale are affected by response set, it does not support the notion that response set is the primary determinant. The lack of consistent positive correlations between the F scale and reversed F scales in the four studies cited therefore *cannot* be attributed primarily to a tendency to acquiesce.

There are other reasons for questioning an attempt to equate responses to the F scale and acquiescent response set (2, 4). The F scale correlates highly with many measures which are free from this artifact, e.g., the analysis of incomplete sentences reported in *The Authoritorian Personality*. (One of the oversights connected with much of the justifiable methodological criticism of that research is that the methodologically more sophisticated work such as this has been largely ignored.) Another qualification should be apparent to anyone who has ever scored the F scale and observed how remarkably inconsistent many people can be in their responses—can a response set underlie such variability? Finally, most of us who have ventured outside the ivied halls have known individuals who emphatically endorse statements such as those embodied in the F scale. Yet they are not only reluctant to acquiesce to statements reflecting approval of any liberal political activity of the past quarter century but emphatically affirm that anyone making such statements is a fool if not a traitor. Among those who represent the quintessence of the F syndrome, placid agreement to explicit liberal statements is rare in our experience.

The preceding considerations indicate that acquiescent response set is one component of scores on the F scale but that its importance has been over-emphasized in the four studies using reversed items. A scrutiny of the data suggests two possible reasons. One is that the reversed items which were used have deficiencies as meaningful psychological opposites to original F-scale items. A second reason appears to lie in the distributions of F-scale scores in the samples studied; these are not congruent with the supposition that the high scoring subjects are acquiescent. These points are worthy of further elaboration.

*Problems in Reversing F-scale Items*

The following considerations seem important to us in the construction of item reversals which should be accepted by those who reject F-scale items and rejected by those who accept them:

1. *Logical opposition to the original items.* This point is by far the simplest and the studies under consideration are not subject to criticism on this basis. We do not, however, consider logical opposition in itself a sufficient condition for meaningful reversals. This point may be illustrated by one of the few items which was reversed in all four studies—"familiarity breeds contempt." The reversals of this laconic aphorism were:

Familiarity does not breed contempt. [2]
Familiarity breeds respect more often than it does contempt. [7]
We are bound to admire and respect a person if we get to know him well. [13]
The more you get to know people the more you get to like them. [17]

These statements are all *logically* incompatible with the original item. Yet they differ in wording and, we would argue, in their *psychological* incompatibility with the original statement and one another.

2. *Avoidance of extremity of wording.* The statements in the F scale are frequently couched in all or none terms: "everybody," "the most important," "never do anything," "no sane, normal, decent person," "no weakness or difficulty," etc. Since this form itself has been hypothesized as appealing to those high on the F scale and who are presumed to be intolerant of ambiguity it would appear appropriate to couch reversals in probabilistic form by using such qualifications as "some," "many," "one of the most important," etc.

In varying degrees, items used in the four studies failed to meet this criterion, e.g.:

*No* people are born with an urge to jump from high places. [2]
It is *completely absurd* to believe that some people may be born with an urge to jump from high places. [7]
It is *known with complete certainty* that the urge to jump from high places is learned, not inborn. [13]
*All* the mysteries surrounding our lives will sooner or later be cleared up through the progress of science. [17][8]

[8] Leavitt, Hax, and Roche did not reverse the "urge to jump" item. The italics are ours.

The theory underlying the F scale suggests that a truly nonauthoritarian individual would be taken aback by the unqualified wording of such statements. Agreement with the underlying premise that suicidal tendencies are most plausibly explained by the vicissitudes of an individual's life circumstances rather than heredity might be masked by a nonauthoritarian disinclination to agree with a dogmatic statement.

3. *Psychological opposition to original items.* Many of the reversals in these studies may be nondiscriminating because they are irrelevant to the *psychological* significance of the originals. In some instances we would predict that the reversals would be accepted by *both* high and low scorers on the F scale; in other cases that the reversals would not be acceptable to either group. Our judgments in these cases are largely subjective since they are based upon our interpretation of the theory underlying *The Authoritarian Personality* and familiarity with subsequent studies. Since it is difficult to formulate general rules in such a situation, we shall cite a few of the many statements that are relevant.

Consider two reversals of the statement: "Nowadays when so many kinds of people move around and mix together so much a person has to protect himself especially carefully against catching infection or disease from them":

Nowadays, since democracy demands that people of widely different backgrounds and station mix together, a person should not be finicky about catching disease from any of them. [13]
You need to get out and rub elbows with all kinds of people to get new ideas and broaden your understanding of life. [17]

The first reversal should presumably be accepted by low scorers since they, theoretically at least, should not be over-anxious about "contamination" by others. High scorers might also be expected to accept the reversal since it is loaded in the direction of a conforming response; rejection of the statement implies rejection of democracy, and, despite their implicitly anti-democratic ideology, their need for conformity is usually associated with the acceptance of the *symbols* of the sanctioned democratic ethos.

The second reversal, it is suggested, misses the psychological point of the original state-

ment which was intended to arouse anxiety about catching diseases from unwashed outgroups. Instead, a folksy democratic statement was used. The basic concept is unlikely to be rejected by low scorers although they might wonder about the "need." Rejection of the item by high scorers would imply that they were intraceptive, which runs counter to theory. For these reasons, the item should theoretically be accepted by both high and low scorers and thus be nondiscriminating.

These considerations suggest that one of the reasons why the scales composed of reversed items fail to correlate with the F scale lies in the nature of many of the items of which they are composed. Many strike us as being subject to rejection by low scorers because of their extremity of wording, irrationality, or lack of compatibility with the psychological significance with items in the F scale. Other reversed items are believed to be nondiscriminating because they might reasonably be expected to be accepted by both high and low scorers. There are enough items subject to these strictures in the studies under review to afford a plausible explanation for the findings reported without the necessity of invoking response set as *the* crucial variable. These considerations are believed to be especially crucial for the studies of Bass, who explicitly attempted to construct *logical* rather than *psychological* reversals, and of Jackson and Messick whose reversals were in the *style* of the original items.

*Acquiescent Response Set and Sampling*

Three of the four samples under review used college samples, apparently largely composed of students at the beginning of their college training. The majority of respondents in the fourth (17) appear to have been highly educated. This point is of methodological interest because college samples do not tend to make *absolutely* high scores on the F scale. More precisely, if the range of item mean scores is defined as ranging from 1.0 (strong rejection of every item in the scale) to 7.0 (strong acceptance of every item), it is found that most college samples reported in the literature have a mean score below the theoretical neutral point of 4.0.

This distribution of scores is important in evaluating the purported relevance of response set to the F scale. Those in the lowest quarter of the typical distribution usually have an item mean score of well below 3.0 which indicates that they must fairly consistently disagree "slightly" or "somewhat" with the items with an occasional agreement or "strong" disagreement. Those who score in the highest quarter typically have mean scores in the 4.0 or 5.0 range with a few scoring slightly higher.

The importance of such distributions for the equation of high scores on the F scale with acquiescence is due to the fact that "neutral" responses are not customarily allowed. Any scores between 3.0 and 5.0 therefore have to be obtained by balancing off agreement and disagreement with specific items. The typical college student "high" authoritarian *cannot*, by virtue of the scale format and scoring, *be consistently accepting items*. This fact suggests that if these studies are based upon typical college samples, the lack of positive correlations reported cannot be attributable to acquiescence response set on the part of the high scorers in these samples since the latter are inconsistent in their responses. (It is hypothetically possible that some of the reported correlations are due to "negativism" on the part of low scorers who reject both the original F-scale items *and* the reversals!)

It is impossible to delineate fully the possible effects of the range of responses on the F scale in the samples used in these four studies. Bass (2) reports an item mean of 2.83 which is equivalent to one of 3.83 in the conventional scoring of the F scale (a constant of three rather than four was added to raw scores). He reports an *SD* of 2.00 (2, Table 2, p. 620). This means, if correctly reported, that his sample was necessarily uniquely bimodal on the F scale since it would be impossible to have a normal distribution on a scale with a six-point range and a standard deviation of a full two points. In view of this and other errors in these data, no reliable estimates can be made about the characteristics of the sample.

Chapman and Campbell do not report the mean or distribution of F scale responses for their two samples. They used a five-point response scale with an undecided category so that the preceding comments about the lack of a neutral category do not apply to their data. An item mean computed from unpublished data indicates that the item mean for their

second sample was 2.90, just below the theoretical neutral point of 3.00. Since these data were in the form of responses to items, it was not possible to compute an *SD* based upon *S*s. It is therefore impossible to determine the extent to which high scoring *S*s were consistent in their acceptance of F-scale items.[4]

Jackson, Messick, and Solley report an item mean of 3.10 with an *SD* of .67 on 42 *S*s. Their scoring procedure differed slightly from the conventional one and yields a mean equivalent to 3.77. If their distribution of scores was normal, approximately one or two *S*s would have scored above 5.0. Leavitt, Hax, and Roche studied two small samples of foremen, (*n*'s = 6 and 13), two of executives (*n*'s = 34 and 52), and one of Unitarians (*n* = 28). A composite plot of their data (17, Fig. 1, p. 219) indicates that low scoring *S*s greatly outnumbered high scoring ones. As nearly as we can determine by inspection, eight *S*s had an item mean agreement varying between "slight" and "somewhat" and a single *S* had an item mean above 6.0. The relatively high scorers in these samples thus fell largely within or slightly above the neutral range.

### *Additional Comments*

It should be emphasized that the preceding comments about these studies are based upon an advantage that the original authors did not have, i.e., the opportunity to be familiar with one another's work. The seemingly paradoxical finding in independent studies that F-scale reversals did not correlate appropriately with the F scale raised questions which led to analysis of possible reasons. Chapman and Campbell have an insightful discussion of the problems of reversing items and because of the difficulties did not try to reverse more than ten items in their initial attempt (7, p. 129). Since they were the only investigators to report an item analysis, a study of their items which did successfully reverse gave us clues as to the possible failures of their other items and those of the other investigators. Further hunches were suggested by Bass' analysis of reversed items as to the degree of judged logical difference from the originals, the strikingly negative correlation found by Jackson *et al.* with reversals that were stylistically similar to F-scale items, and the different correlations among samples reported by Leavitt *et al.*

## AN EXPLORATORY INVESTIGATION OF REVERSED F-SCALE ITEMS

The present study was undertaken to explore the implications of our examination of earlier studies. Clarification of the problem of the role of response set in the F scale required, we believed: (*a*) reversed items which were constructed according to the criteria previously specified and (*b*) the testing of these items on samples believed to contain a wide range of scorers on the F scale.

### *The Item Pool*

It was decided that the larger the number of reversals attempted, the greater the probability of empirically finding ones which were correlated with the F scale. In two studies (7, 17), the researchers did not test reversals of all the items in the F scale. In three of the studies (2, 13, 17), F-scale items were reversed although they had been explicitly mentioned in *The Authoritarian Personality* as being poor discriminators (1, Nos. 22 and 44, pp. 254–257). We attempted to reverse every F-scale item which was acceptable in its original form.

The final version of the F scale described in *The Authoritarian Personality* contained 29 items. We attempted to reverse the 27 which were found to be discriminating. One, "wars and social troubles may someday be ended by an earthquake or flood that will destroy the whole world," baffled us in our attempts to construct a meaningful reversal. Three items which were in the subsequent Form 60 A were added to the pool. A fourth additional item, taken from the earlier Form 60, was also

[4] The item mean is higher than generally found in schools of Northwestern's caliber. Since Chapman and Campbell did not randomly sample F-scale items and used but 16 of them, the possibility of bias in the items was considered. The 14 items taken from the Form 40 of the F scale had an item mean (our computations) identical with the other items in this form in the original samples (based upon 1, Table 9 (VII), p. 260). Their other two items were taken from the Form 60 and had an item mean of 4.46 in contrast to that of 3.71 on the whole scale (1, Table 6 (VII), p. 253). Since there is reason to believe that the relative "pull" of F-scale items is fairly consistent in college samples, a slight contribution to the unexpectedly high item mean in this sample may be attributed to the popularity of these two items. The effect of allowing a "neutral" option upon item mean scores is unknown.

added. It had discriminated between high and low scorers (1, Table 6, (VII), item 41, p. 258) but, for unstated reasons, was not included in subsequent revisions of the F scale.

*The Item Reversals*

After discussion of the problems of item reversal along the lines indicated above, the authors individually attempted to write one or more reversals of each item in the pool. These reversals were then discussed and one of them, or more often a modification of one, was accepted for provisional inclusion in the reversed scale. The reversals were then circulated among colleagues familiar with the F scale; their critical comments were solicited and, in most cases, accepted.

*The Subscales*

Two questionnaires (Forms A and B) were prepared. Each was composed of 15 of the 30 original F-scale items and reversals of the other 15 since it was believed that the presence of the original and reversed form of the same item in the same questionnaire would lead to possible artifacts in response. The assignment of items to the two forms was based upon the use of a table of random numbers and the serial numbering of the items within the questionnaire was also random. There was no a priori reason for assigning items since earlier research had indicated that individual items in the F scale fell into different empirical clusters in different samples (8). Each scale was a mirror image of the other, i.e., if a given item appeared in its original wording on Form A its reversal was in the same serial position on Form B. The item reversals used, their allocation to Form A or B, their position within the subscales, and the source of the original item are given in Table 1.

*Scoring*

The customary procedure in scoring the F scale was followed, respondents being given an option of agreeing strongly, somewhat, or slightly or disagreeing slightly, somewhat, or strongly. No neutral or undecided option was allowed, but omission of an answer was scored as neutral. Both scales were scored so that high scores reflected "authoritarianism" as indicated in Table 2.

*The Samples*

We were interested in obtaining samples which included *S*s who would show great variability of scores on the F scale and were especially interested in finding high scorers. Like other investigators, we were hampered by the difficulty of obtaining noncollege samples. Through the cooperation of friends, the scales were administered to their classes at two colleges that were believed to contain a higher proportion of high scorers than would be found in most colleges with which we were familiar. One of these was a private university in the Southwest, the other, a Northeastern public university. The scales were administered during the 1956 summer session when the student bodies were more heterogeneous than usual since they also contained many elementary and high school teachers who were attending summer school.

*S*s in each of the two college samples were divided in half by administering odd (Form A) and even (Form B) scales to adjacent *S*s. Results on four samples will be discussed: SW-even ($n = 75$), SW-odd ($n = 75$), NE-even ($n = 39$), and NE-odd ($n = 40$). Item mean scores and *SD*s on the 15-item F subscales were respectively: 3.81, .97; 3.39, .93; 3.26, .99; 3.35, .93. We are unable to explain why the SW-odd sample scored so surprisingly much lower than the SW-even sample since the alternate forms were distributed to adjacent students and the two SW samples do not differ significantly in proportion of males (roughly two-thirds), age, or major in college. Roughly half of the NE samples were male. Fourteen per cent of both samples were psychology majors, 25 per cent were education majors, and 19 per cent were business adminstration majors. The bulk of the remainder of the NE samples was composed of nursing students, that of the SW samples of students in engineering and liberal arts. Respondents ranged from freshmen to graduate students in all samples.

## RESULTS

*Item Analysis*

In analyzing the reversals, scores made by the *S*s on the appropriate F subscale were divided into quarters. Two comparisons were then made of the scores on individual reversed items—the magnitude and direction of the individual item mean differences of those in

TABLE 1
REVERSALS OF F-SCALE ITEMS

| Orig. No.[a] | Form | No. | Reversed Statement |
|---|---|---|---|
| 31 | B | 1. | There are times when it is necessary to probe into even the most personal and private matters. |
| 37 | A | 2. | It would be a good thing if people spent more time thinking and talking about ideas just for the fun of it. |
| 21 | A | 3. | If it weren't for the rebellious ideas of youth there would be less progress in the world. |
| 9 | B | 4. | When a person has a problem or worry, it is best to face it and try to think it through, even if it is so upsetting that it keeps him from concentrating on other things. |
| 19 | B | 5. | Insults to our honor are not always important enough to bother about. |
| 2 | B | 6. | There are many difficulties a person cannot overcome no matter how much will power he has. |
| 39 | B | 7. | It's nobody's business if someone is a homosexual as long as he doesn't harm other people. |
| 8 | A | 8. | It's all right for people to raise questions about even the most sacred matters. |
| 29 | B | 9. | It is highly unlikely that astrology will ever be able to explain anything. |
| 13 | A | 10. | In the long run it is better for our country if young people are allowed a great deal of personal freedom and are not strictly disciplined. |
| 18 | B | 11. | Even though people of all sorts mix together nowadays, you don't have to worry very much about catching an infection or disease. |
| 34 | B | 12. | Even if there were no immoral, crooked, or feeble-minded people there would still be many serious social problems. |
| 16 | A | 13. | An urge to jump from high places is probably the result of unhappy personal experiences rather than something inborn. |
| 26 | A | 14. | It doesn't make much sense to divide people into groups like the weak and the strong; too many people are strong in some ways and weak in others. |
| 4 | A | 15. | The findings of science may some day show that many of our most cherished beliefs are wrong. |
| 41 | A | 16. | People ought to pay more attention to new ideas, even if they seem to go against the American way of life. |
| 43 | B | 17. | You may dislike a person very much, but the chances are that if you get to know him well you'll have more respect for him. |
| 1 | A | 18. | One of the most important things children should learn is when to disobey authorities. |
| 59 | A | 19. | Most people can be trusted. |
| 53 | B | 20. | You'll do better work in the long run if you do things your own way rather than by following tried and true procedures. |
| 23 | B | 21. | It is more important for this country to have a just legal and political system than a series of trustworthy leaders, however courageous, tireless, and devoted they might be. |
| 25 | A | 22. | Sex crimes, such as rape and attacks on children, are signs of mental illness; such people belong in hospitals rather than in prison. |
| 38 | A | 23. | Many people have too great a fear of plots hatched in secret by politicians. |
| 42 | A | 24. | It's only natural for people to sometimes have thoughts about hurting a close friend or relative. |
| 41 | B | 25. | The artist and professor are probably more important to society than the businessman or manufacturer. |
| 27 | A | 26. | Most honest people admit to themselves that they have sometimes hated their parents. |
| 6 | B | 27. | Human nature doesn't make war inevitable; man may some day establish a peaceful world. |
| 35 | B | 28. | In spite of what you read about the wild sex life of people in important places, the real story is about the same in any group of people. |
| 12 | A | 29. | People should be willing to overlook failures in the manners and unpleasant personal habits in other people. |
| 58 | B | 30. | Books and movies ought to give a more realistic picture of life even if they show that evil sometimes triumphs over good. |

[a] All items are taken from 1, Table 7 (VII), pp. 255–257 except for No. 16 which was found in Table 4 (VII), p. 250 and Nos. 19, 20, and 30, the originals of which are in 8, Table 3, p. 464.

the upper and lower quarters on the F subscales and the pattern of differences in all four quarters.

Of the 30 reversals, 22 discriminated in the appropriate direction between high and low quarters on the F scale in both samples responding to them, four in one sample but not the other (Nos. 7, 11, 14, and 17), and four discriminated negatively in both samples responding to them (Nos. 1, 20, 21, and 23). This differs from chance expectancy at better than the .01 level as tested by chi square. The attempt to find workable reversals is thus viewed as moderately successful.

Individual F-scale items were similarly analyzed against extreme quarters on the total scale. All discriminated in the appropriate direction but there was great variability in discriminatory power and mean level of acceptance.

### *An Examination of Response Set*

Although these findings indicate that F scale items are, at least within limits, reversible,

they do not clarify the more basic problem as to how response set is linked with the F scale. Given the fact that the F scale has been found to be correlated with a host of measures that do not involve response set, that its construct validity is tenuous, and that response set plays some role—although not the dominant one in the acceptance or rejection of its items—how best to proceed?

Since we had selected our reversed items against the criterion of scores on F subscales we decided to examine further the most successful five in each form. This procedure eliminated the negatively discriminating items and those of marginal discrimination and left only those found to be systematically related to response to the F scale.

We also pared down the number of F-scale items for more intensive scrutiny to five in each subscale. These were the ones with the highest discriminatory power in the item analysis. So doing tended to get rid of not only those items which were poor discriminators but those which were widely rejected or accepted. For example, the item mean on the statement, "People can be divided into two distinct classes: the weak and the strong," was but 1.77 in the SW-even sample and even those scoring in the high quarter tended to reject this "authoritarian" item, their item mean score being but 2.0 (disagree "somewhat"). On the other hand, the SW-odd sample tended to accept the statement, "The wild sex life of the old Greeks and Romans was tame compared to some of the goings-on in this country, even in places where people might least expect it," the item mean being 5.15. Slightly greater acceptance than rejection of this statement on the part of low scorers was found (item mean of 4.20). The ten reversed items (five from each form) used in this analysis, were Nos. 3, 5, 13, 15, 16, 18, 25, 28, and 30 in Table 1. The F-scale items were the originals of Nos. 1, 6, 8, 9, 15, 18, 20, 22, 24, and 30.

These short five-item scales had a mean part-whole correlation of .82 with the longer 15-item forms and thus may be viewed as tapping the essence of the F scale. They are also useful since they are presumably the most susceptible to response set if response set is part and parcel of the "authoritarian" syndrome. After all, items that are universally accepted or rejected cannot yield much information about differential response set among individuals.

TABLE 2
SCORING OF STRAIGHT AND REVERSED F-SCALE ITEMS

| Response | Score | |
|---|---|---|
| | Straight F | Reversed F |
| Agree Strongly | 7 | 1 |
| Somewhat | 6 | 2 |
| A Little | 5 | 3 |
| No answer | 4 | 4 |
| Disagree a Little | 3 | 5 |
| Somewhat | 2 | 6 |
| Strongly | 1 | 7 |

TABLE 3
CORRELATIONS BETWEEN LONG AND SHORT FORMS OF STRAIGHT AND REVERSED F-SCALE ITEMS

| Sample | Scales Being Correlated | | |
|---|---|---|---|
| | 15 item F+ and F− | 15 item F+ and 5 item F− | 5 item F+ and 5 item F− |
| SW-odd | .35 | .42 | .45 |
| SW-even | .14 | .27 | .23 |
| NE-odd | .34 | .27 | .37 |
| NE-even | .45 | .66 | .44 |

The correlations between these abbreviated scales and the original 15-item scales is given in Table 3. Since the items in the short scales were selected against the longer F subscales, it is not surprising that the magnitude of the correlations is similar. It is clear, however, that the correlations fall far from what would be expected if the reversals were adequate and response set did not play some role.

Scores of the 299 *Ss* in the four samples were grouped and ordered in terms of their item mean response on the F scale. The 58 *Ss* making scores of 5.0 and over were considered to be high scorers and their responses to the reversed items were examined. Sixteen of these individuals had item mean scores of 6.0 or higher on the short F scale. Despite this high degree of "acquiescence" with "authoritarian" items, it did not generalize to the reversed items. Their scores here all fell in the neutral range (3.2 to 4.8) with the exception of one slightly deviant case with a score of 5.0 (disagree slightly). There are many ways of balancing off responses to attain a "neutral" score and these *Ss* displayed a most consistent inconsistency in doing so. They used the extreme categories in answering the reversed items; 84 per cent of their total responses were in terms of "somewhat" or "strong" acceptance *or* rejection. No one accepted or rejected more than three or less than two of the reversed items. Further, there was no observable difference in the acceptance or rejection of specific items.

It is clear that an acquiescent response set does not account for these results. These individuals who wholeheartedly and consistently endorsed the most discriminating F-scale items displayed an enthusiastically random pattern of response to the reversed items. Our interpretation is that they have a tendency to make extreme responses and are ideologically somewhat confused.

Those scoring from 5.0 to 5.8 ("slight" to almost "somewhat" agreement) on the short F scale differ in showing extreme variability on the reversals, the item mean range varying from 1.8 to 5.6. Those at the extremes of this continuum display quite different behavior.

Five of these individuals had item mean scores of over 5.0 on both the F and reversed F subscales, a higher score on the reversed scale, incidently, than any of 16 individuals with scores of 6.0 or over on the F subscales considered earlier. These are clearly ideologically consistent individuals and are moderately fascistically oriented. Their modal response pattern is that of accepting four out of five items on the F subscales and rejecting four out of five on the reversed F subscales. The small *n* does not suffice for an evaluation of the intensity of response although there is a lack of use of extreme categories (there was not a single "strong" approval of a reversed item). This fraction of the *S*s shows no evidence of response set.

Acquiescent response set is definitely manifested in a group of 13 *S*s with F subscale scores above 5.0 and reversed subscale scores of 3.0 or below. The modal pattern was acceptance of four of five items on *both* the F and reversed F subscales. The "authoritarianism" manifested by these individuals on the F scale appears nonideological in nature and a reflection of a tendency to agree with incompatible items.

The remaining 24 *S*s in this range of scorers show a mixed pattern of responses, the mean on the reversed F subscales falling between 3.2 and 4.8. Although tendencies in the response set direction may be present among some of them, it is equally plausible that they tended toward a consistency which was masked by the analysis of a relatively small number of items.

A similar analysis was made of those scoring low on the F subscale. There were 21 *S*s who disagreed "somewhat" or greater (scores of 2.0 or below). Thirteen of these had a general pattern of acceptance on the reversed items and may be considered consistent lows. Eight, however, had mean reversed scores in the neutral range and thus resemble mirror images of our extreme high scorers. They differ from the latter, however, in not predominantly utilizing extreme responses to items: 38, 40, and 23 per cent of their responses fell in the "slight," "somewhat," and "strong" agree and disagree categories. There is thus no evidence of response set among these respondents.

The moderately low scoring *S*s who scored from 2.2 to 3.0 on the F subscales can be identified as ideologically consistent or as subject to response set just as those with moderately high scores. A total of three *S*s displayed a consistent *negative* response, i.e., they tended to disagree with items on *both* subscales. Eighteen *S*s were identified as consistently low since they accepted the reversed items while rejecting the original ones. The remaining 20 *S*s in this range had scores in the neutral range on the reversed scales and are considered a mixed group.

Within these samples, at least, three general findings emerge: (*a*) there is a much higher proportion of ideological consistency among low than high scorers, (*b*) response set is more prevalent among those who moderately agree than those who agree extremely and although slightly present among those who moderately disagree is not found among extreme disagreers, and (*c*) there is a higher proportion of ideological inconsistency on the reversed items among extreme high scorers than among other groupings.

The consistency of the low scorers is congruent with the findings of Leavitt *et al.* They combined their groups and correlated the original F-scale items used and their reversed scale and found a positive correlation in the bottom third of the combined samples of +.45 (the high third on the F scale had a correlation on the two scales in the *negative* direction: −.32). Although we have not replicated their analysis, an inspection of our data indicates (as can be easily guessed from the earlier discussion of response patterns) that the low scorers are responsible for the overall positive correlations; the correlation between original and reversed items among the high scorers in these samples would not differ appreciably from zero.

The greater presence of response set among those with moderate rather than extreme attitudes is not at all surprising. This is precisely what would be expected according to Cronbach's formulation of the conditions under which response set is most likely to be found. A person sophisticated enough to reject most of the F-scale items out of hand has a well enough formulated ideology to know where he stands and is unlikely to find reversed items vague or unfamiliar. An extreme high scorer is of different persuasion but his devotion to the irrational does not indicate bland acceptance of contrary statements, at least in light of current theory and knowledge. Although the preceding point seems exceedingly obvious, it has not been prominent in recent discussions of response set. One possible value of the present findings, limited as they may be, may be the reiteration of basic considerations about the relationship between extremity and intensity of attitudes. These, in conjunction with Cronbach's observations, suggest that it is precisely those persons who do not have strong attitudes pro or con who should be most apt to display response set.

The one unexpected finding was the complete abandon with which the extreme high scorers rejected or accepted the reversed items. There is considerable discussion in *The Authoritarian Personality* about the irrationality of such individuals but no indication that it would take such a remarkable form.

## FURTHER STUDIES

Although the preceding findings appeared plausible, they were descriptive and left several matters in doubt. First, how stable were they? Our procedure of concurrently analyzing items from separate samples is similar but not identical to doing an item analysis on an original sample and cross-validating on a new one. Further, our results were obtained on samples of summer school students who can be viewed as representative of neither the more usual college samples nor of the general population. Replication of the findings in other samples therefore appeared called for.

The second matter of concern was the unanticipated behavior of the extreme high scorers on the reversed items. Some confusion would not have been surprising but the failure to find a single high scorer who would firmly reject reversed items was most puzzling. Since these individuals were key members of our samples whose responses served as the basis for the item analysis, it was not clear whether this might be due to an artifact arising from the method of item selection, whether it reflected the possibly aberrant nature of our samples, or whether it was indeed a characteristic of high scorers on the F scale.

Data have therefore been collected on eight additional samples. These, in all candidness, were "targets of opportunity," i.e., they were not systematically selected to test specific hypotheses about the relationship between the straight and reversed forms of the F scale. They are samples made available informally through the kindness of friends in testing their classes or in permitting the inclusion of the scales in ongoing studies directed toward other ends.

### *The Scales*

The reversed F scale was composed of Items 3, 9, 13, 15, 16, 18, 25, 26, 28, and 30 in Table 1. With the exception of Item 26, these are identical to those used in the subscale analysis in the exploratory study.

The F-scale items were the originals of Nos. 1, 6, 8, 10, 12, 20, 22, 23, 24, and 27 in Table 1. They differ from those used in the subscale analysis in that the originals of Items 9, 15, 18, and 30 were deleted and substitutions were made to prevent the presence of a reversal and its original in the same questionnaire.

Ordering of the items was random. The customary response options were provided. All samples received the same 20-item scale under conditions noted in the description of each.

### *The Samples*

1. Undergraduate students in a Northeastern public university were given the 20-item scale during the first week of a course in introductory psychology. It was administered by the instructor (male) in charge of the classes with assurances about the preservation of anonymity and the use of the results for research. The respondents were young (median age of 18.7 years), 72 per cent were male, and they were predominantly lower middle-class in background. Sixty-two per cent were freshmen and 30 per cent were sophomores.

2. Students in a graduate seminar in social psychology at a major Eastern university filled out the 20-item scale and also were required as part of the course to administer it to friends and acquaintances. Protocols of individuals familiar with the F scale were discarded. The sample as a whole did not respond to the questionnaire under standardized conditions. It is unusually high in years of formal education.

3. Students in a summer school course in remedial reading at a prominent Eastern teachers college filled out the 20-item scale at the end of a class period. It was administered by a female introduced by the instructor as carrying out research. Most of the respondents were female school teachers from a broad geographic area and had a median of roughly 15 years of teaching experience. Most of them had done their undergraduate work in teachers colleges and were taking the course to maintain teaching credentials.

4. A nonrandom but fairly representative sample of registered Washington lobbyists was given three separate questionnaires to be filled out and returned after each had been interviewed about his role as a lobbyist. One of the questionnaires contained six F-scale items and four reversals. These were interspersed with 25 other items constructed by the

senior author which were largely taken from the writings of Machiavelli. A five-point response scale which included a neutral or undecided category was used. The majority of this sample held law degrees and most had previously held elective or appointive political office.[5]

5–8. These samples consisted of senior medical school students (with one junior class) in three top-ranking medical schools. Students in two which were located in the East had predominantly received their undergraduate training at the more prestigeful Eastern liberal arts colleges. Those in a Midwestern school were largely from the better Midwestern colleges.

Items from the 20-item scale were randomly distributed among 20 Machiavellian items (half of which were worded in the reversed direction) in an "opinion inventory." These and other psychological measures were part of a lengthy questionnaire devoted to experiences in medical school. In some instances the questionnaire was answered in a class setting, in other cases individuals filled it out at home.[6]

*Results*

The data summarized in Table 4 indicate the consistency with which positive correlations between the F and reversed F scale were found. The samples varied in education from beginning college students to those with eight years of undergraduate and professional experience in the finest of our educational institutions, from predominantly female elementary and high school teachers to lobbyists with a bad case of "Potomac Fever," some answered the questionnaires anonymously in class while others filled them out in conjunction with personal interviews, the items were sometimes presented as an individual scale and were sometimes mixed with other items and in five samples were given along with a large number of unrelated questions, and in general the customary seven-point scoring system was used, but in one case five-point response categories with a neutral option were used.

[5] This study of the role and personality attributes of lobbyists was carried out under the auspices of the Brookings Institution under the supervision of Lester Milbrath.

[6] The samples were studied as part of ongoing research on the Sociology of Medical Education by the Bureau of Applied Social Research at Columbia University under a grant from the Commonwealth Fund.

Despite all these variations that might possibly affect the nature of the responses positive correlations were found in all eight samples. When these are combined with the findings in the exploratory study it is found that in twelve of twelve samples using various combinations of these items the results are such as to indicate that response set is not important enough to produce negative correlations.

The other problem that intrigued us—that of the behavior of the extreme highs on reversed items—is partially clarified by a closer inspection of these data. The distribution of responses on the F scale was leptokurtic in the undergraduate sample, there being but eleven individuals with a mean score of 3.0 or below and five with scores of 6.0 or above. The heavy concentration of respondents in the neutral and slightly accepting range casts some light on the very low correlation between the scales in this sample since the generalizations made earlier about the relationship between mean acceptance of F-scale items and acquiescence were confirmed. However, of the five high scorers, two displayed acquiescent tendencies, two behaved similarly to the extreme highs discussed earlier, and one was consistent.

Further evidence that the extreme high was not necessarily inconsistent came most strongly from the lobbyist sample where there was very little evidence of acquiescence. Indeed, we found one "perfect" high scorer (within the limits of the items used) who strongly accepted all six F-scale items and strongly rejected all four reversals. Unfortunately, there were not enough extreme high scorers among these samples to serve as a sound basis for generalizations.

The wide range of reliabilities on both the original and reversed F scales and the varying correlations between them posed problems in interpretation. It should be added that where data from other attitude scales were available, they indicated that the amount of ideological consistency present was not unique to the F scale. The Washington lobbyists, for example, had the highest internal consistency on the Machiavelli scale of nine samples totaling some 2,000 *Ss* tested to date. This finding parallels their results on the F scale. Similarly, Sample 5, composed of a particularly aberrant (statistically speaking) group of medical students, showed inconsistency similar to that on the

TABLE 4

SUMMARY RESULTS ON F AND REVERSED F IN EIGHT SAMPLES

| Sample | N | F Scale $\overline{M}$ | F Scale $\overline{SD}$ | F Scale Rel. | Reversed F Scale $\overline{M}$ | Reversed F Scale $\overline{SD}$ | Reversed F Scale Rel. | Correlation between F and Reversed F Scales[a] |
|---|---|---|---|---|---|---|---|---|
| 1. Northeastern undergraduates | 152 | 4.54 | .84 | .52 | 3.73 | .78 | .43 | .10 |
| 2. Graduate students | 57 | 2.74 | .95 | .78 | 2.72 | .93 | .57 | .58 |
| 3. Lobbyists | 86 | 3.05[b] | .59 | .54[c] | 3.03[b] | .81 | .41[d] | .55 |
| 4. Remedial reading teachers | 61 | 3.43 | 1.05 | .68 | 3.20 | .94 | .60 | .29 |
| 5. Midwestern medical students (seniors) | 72 | 3.50 | .69 | .34 | 3.35 | .55 | .18 | .45 |
| 6. Midwestern medical students (juniors) | 80 | 3.55 | .88 | .66 | 3.47 | .70 | .26 | .23 |
| 7. Eastern medical students[1] (seniors) | 73 | 3.38 | .92 | .75 | 3.35 | .91 | .72 | .38 |
| 8. Eastern medical students[2] (seniors) | 121 | 3.76 | .86 | .58 | 3.17 | .68 | .42 | .48 |

[a] All correlations were corrected for attenuation.
[b] A five point scale was used in this sample. These figures are comparable to slightly over 4.0 in other samples.
[c] Based on only six items.
[d] Based on only four items.

F scale when responding to Machiavellian items but were very consistent on a measure of manifest anxiety. Such findings suggest that samples vary widely in the amount of acquiescence that they will display in response to items of an ideological nature.

The investigation of differences from sample to sample demands a more refined technique of measuring response set than that used in the preceding examples. Before pursuing the substantive problem, it is necessary to indicate the method used.

## A METHOD OF MEASURING ACQUIESCENCE

The diagnosis of acquiescence used with the four exploratory samples was crude and obscured the interaction of content and response style. The correlations and reliabilities indicated in Table 4 suggest widely varying probabilities of response set in different samples. These respondents were subjected to different numbers of items and in some cases different choices of response categories. In considering the problem of comparability from sample to sample, it was decided that it was simplest to use the number of statements agreed to on the F and reversed F scales as the basic data. This convention permitted the use of the binomial expansion and the ascertainment of the probability of a given pattern of response.

Although the technique is basically simple, we are unfamiliar with examples of its application and are illustrating its use with two five-item scales.

There are six possible degrees of agreement for each five-item scale. Cross-multiplication of the binomial frequency for each response category on each scale indicates the chance probability in 1024 times of the frequency of a given cell in the matrix, e.g., the probability of a given individual agreeing with all five items in each scale is one in 1024. The expected frequency of any given pattern of response is indicated in Fig. 1.

The second information in the matrix is the discrepancy scores between the two scales. If the two scales are measuring essentially the same thing—and the means on two sophisticated samples (graduate students and lobbyists) as indicated in Table 4 suggest that this is approximated—then an individual who agrees with all five F-scale items should tend to disagree with all five reversed items. A simple way to check for consistency is to reverse the number of agreements on the reversed scale so that agreement with all five is scored 0 and disagreement with all is scored 5. Subtraction of the reversed scale score from the F-scale score gives a measure of content consistency. A perfectly consistent individual should have identical scores on both scales. A person who is indiscriminately accepting will have a discrepancy score of +5. If he is recalcitrantly negativistic he will have a discrepancy score of −5.

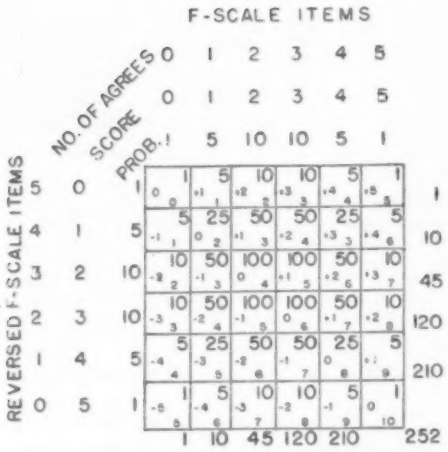


FIG. 1. MATRIX FOR DETERMINING THE PROBABILITY OF A GIVEN PATTERN OF RESPONSE ON 5 ITEMS F AND REVERSED F SCALE

The number in the upper right hand corner of each cell indicates the chance expectancy in 1024 times of a given pattern. The upper small number indicates the discrepancy score; the lower one the total F score on the combined scales. The numbers at the right and bottom are the summed frequencies for the cells with identical discrepancy scores along the diagnonals of the matrix.

The next step is to determine the probability of occurrence of a given discrepancy score. Summing along the diagonals composed of cells with the same discrepancy scores gives the appropriate frequencies. It will be noted that this distribution is the binomial expansion of $2^{10}$. The probability of occurrence of a discrepancy score of +3 or greater is .055 ([1 + 10 + 45]/1024). This is as close to the more customary .05 level as can be obtained and is sufficient for present purposes. Since the matrix is symmetric, discrepancy scores of −3 or greater indicate a tendency to disagree significant at the same level.

A combination of scores on both scales can also be used to obtain a measure of overall F by scoring along the orthogonal diagonal. An individual who rejects all five of the original items and accepts all five reversals is clearly low on the dimension and is given a score of 0 as indicated in Fig. 1. His opposite number obtains a score of 10. Scores of 2 or less and 8 or more differ significantly from chance expectancy.

*Acquiescence in the College Trained Population*

A clearer picture of the operation of acquiescence response set may be had by introducing data collected on one further sample. This is a subsample of the 1956 post election survey conducted by the Survey Research Center of the University of Michigan. Adult respondents selected in a nationwide probability sample were asked their education and 103 reported having at least some college training. In the course of an interview they responded to five F-scale items and five reversed items.[7] The preceding method of analysis was employed and it was found that five per cent were significantly acquiescent, 28 per cent slightly so, 23 per cent were completely consistent, 33 per cent showed slight negative tendencies, and 11 per cent had a significant negative response pattern. It is of considerable interest, in light of current equation of acquiescence with response set, that this sample displayed slightly more negativism than acceptance in responding to items. The amount of observed response set does not, however, differ from chance expectancy.

This finding takes on possible significance since graduate students and Washington lobbyists also displayed a negligible amount of response set. These are samples which are presumably the least confused ideologically. The sample which is most distinctive is that composed of Northeastern undergraduates. A technique similar to that indicated in Fig. 1 was employed for the two ten-item scales they answered. The probability of having a discrepancy score of ±4 or greater is .059.

Thirty-five or 23 per cent of these students showed a significant amount of acquiescence and five (three per cent) showed significant negative tendencies. Although the definition of significance level is slightly different for these undergraduates and the nationwide sample, the difference is quite striking. Why should undergraduates be so much more

[7] The relationship between education and response set in the entire sample is currently being analyzed and will be reported in a separate publication. The items were slightly modified from the wording indicated in Table 1 for use with respondents of varying verbal sophistication. Details of the sampling procedures and the analysis of the interview material will be found in the forthcoming book on the 1956 presidential election by Angus Campbell and Warren Miller. This study was carried out by the Survey Research Center with a grant from the Rockefeller Foundation.

acquiescent than a nationwide sample of adults exposed to college (only two-fifths of whom were graduates)? A precise answer would require a detailed investigation of factors influencing changes over time. It should be noted, however, that such a finding is congruent with a variety of data on attitude changes in the first years of college. Jacob has summarized a number of studies dealing with changes on the F scale and other measures such as favorableness toward civil liberties and other attitudes indicative of liberalism (16). The general finding is that students are most conservative when they enter college. Exposure to the more liberal norms of college usually leads to high scorers dropping out of school, changing in a more liberal direction, or adamantly holding to their original beliefs (Newcomb has furnished graphic illustrations of the last two reactions [21]). The weight of the evidence indicates that this conflict of values is most severe during the initial year or so of college.

Our tentative interpretation is that many among these undergraduates—who were in the very early days of college—were in a stage of value transition and possibly conflict. They still tended to accept F-scale items which are more widely accepted in noncollege circles such as those from which they had just come. At the same time, the reversed items were presumably more in keeping with the value climate of the university to which they were adjusting. Many of the students tended to accept both sets of items. This "acquiescent response set" does not strike us as a mere mechanical way of responding; it probably reflects another facet of Cronbach's discussion of the ambiguity of an item. In this case, items reflecting contradictory views of the world are accepted because the student has not yet decided for himself which is *the* right response; both are supported by different reference groups.

## DISCUSSION

Although the preceding digression is slightly speculative, it has important implications if valid. An initial question about some of the studies emphasizing response set focused upon the actual scale scores of the *S*s and the fact that "high" scorers appeared to be only relatively high and were in actuality inconsistent in their responses. Upon the basis of these later considerations, based, it is true, upon a limited sampling of undergraduates, it would appear that the majority of these studies have used subjects who were most liable to display response set of any populations with exposure to higher education, i.e., undergraduates at an early stage of training.

There is another obvious implication of the relatively high number of acquiescent respondents in the undergraduate sample. Most of the studies involving *S*s high and low on the F scale use samples drawn from undergraduate populations and most usually those in the first two years of college. This means, quite bluntly, that the high scorers are quite likely a mixed group. Some are genuine high scorers with a consistent set of authoritarian attitudes that involve rejection of reversals as well as acceptance of the F-scale items. Others may be simply agreeing to statements of this sort and the fact that some of them are "authoritarian" and others are not is a reflection of their value confusion and not necessarily indicative of potentially fascist inclinations at the personality level.

The possible confusion that might result from treating all individuals with high F-scale scores as equivalent in a study relating the F scale to a measure of intolerance of ambiguity, for example, should be obvious. Those who are high scorers by virtue of being acquiescent are by definition tolerating a considerable amount of ambiguity since they are agreeing to psychologically incompatible statements! That undue reliance on the F scale alone as a measure of authoritarianism might be misleading can be indicated by examining the pattern of responses of those undergraduates in the Northeastern sample who accepted eight or more of the ten F-scale items. Twenty of the 37 turn out on inspection to be significantly acquiescent but not significantly high on the overall measure of F, eleven are significantly authoritarian but not acquiescent, and seven are not significantly high on either measure (one individual with a score of 15 on F and a + 4 discrepancy score was both significantly authoritarian and significantly acquiescent). A virtue of the suggested method of scoring is that it gives two scores to an individual—one a measure of authoritarianism in which response set is balanced out and one a measure of response set. With a large enough number of items, it is possible to have overlap on the two

measures, e.g., two *S*s among the 72 per cent of the graduate student sample who were significantly low on the overall measure of authoritarianism were also significantly acquiescent and another two low scorers were significantly negativistic.

It is not news, of course, to point out that the F scale is not unidimensional. The present argument suggests, however, that there are possibly greater differences among high scorers than low scorers in the customary undergraduate samples. When relatively high and low scorers are compared differences will be found when the low scoring *S*s are different from all other groups. This may account for many of the correlations found with other attitude scales when the generally more sophisticated low scorers differ from both the consistent high scorers and the confused acquiescent high scorers. The generally murky picture that prevails, however, as to the relationship between the F scale and measures of such behavioral tendencies as intolerance of ambiguity and rigidity might arise from the fact that the low scorers are similar to consistent high scorers in having an organized *Weltanschauung* but are more similar to acquiescent high scorers in being able to tolerate certain kinds of ambiguity.

These are speculations and their validity can be determined only by empirical studies using techniques that clarify the meaning of the F-scale scores of a given *S*. It must be emphasized that the majority of the samples reported on in this discussion were accidentally and not randomly drawn so the only generalization about the distribution of response set which is justified is to the nationwide population of adults with some college training where no significant amount of acquiescence was found. The presence of some acquiescence among college undergraduates raises questions about the meaning of some of the results found with other samples of undergraduates but does not necessarily imply that they too are laden with acquiescent individuals.

It should also be noted that the possibility of response set in samples studied by means of the F scale does not necessarily indicate the relevance of response set to studies where the E scale was used as the independent variable. A study on response set by Shelley yielded a correlation of +.27 ($p = .05$) between the anti-Semitism subscale and the use of "like much" responses to 60 pictures on the Perceptual Reaction Test of Berg and Hunt (25). This is a lower correlation than the majority of those cited earlier between the F scale and various measures of response set. Prentice found 16 favorably worded items toward Jews and Negroes correlated less highly (+.78) than did 16 unfavorable statements which correlated +1.00 with the E scale. Although we are in basic agreement with Prentice's conclusion that the use of positive or negative items is of little relevance in the measure of ethnic prejudice, we note as do Gage *et al.* that it does make some difference, a conclusion in agreement with the earlier research of Campbell and McCandless (23). Actually, this slight difference is somewhat surprising since some of the items used by Prentice violate standards we have suggested earlier, e.g., "*No* Jews are especially loud, noisy and overbearing" (23, Footnote 3, p. 420, italics added).

Our interpretation is that in these two studies, slight evidence of response set was found but, given the difficulties of comparing studies using different techniques and samples, that the amount of response set is even less than that found in studies using the F scale. This result is consistent with the general thesis that response set is greatest in ambiguous situations; the use of a minority group as a referent in an attitudinal statement is presumably less vague for most respondents than people in general or the sex life of the ancient Greeks and Romans.

In concluding, the data indicate that acquiescence is not identical with authoritarianism. This interpretation is at variance with the earlier studies reviewed, a discrepancy here attributed at least partially to the fact that many of the reversed items used in them were irrelevant to a meaningful measure of authoritarianism. It also appears that the nature of the samples used was a contributing factor since the high scoring subjects in many of these studies were inconsistent in their responses partly because they were drawn from populations containing a relatively large proportion of individuals who were, although probably but temporarily, ideologically confused.

## SUMMARY

Four recent studies which have found a lack of positive correlations between the F scale

and scales of reversed items were examined. It was argued that the attempts at reversals were often characterized by items that were not psychologically opposed to the original items. It also appeared that calling upon the operation of an acquiescent response set as an explanation of the results was questionable since many of the "high" scorers in these samples were inconsistent in their responses and were not generally agreeing with F-scale items.

A new set of items designed to overcome some of our objections was developed. An item analysis indicated that the majority of the reversed items discriminated positively between individuals high and low on the F scale in two pairs of subsamples. An examination of the response patterns on the most discriminating items indicated that a minority of the respondents showed response set and that they were among the individuals who showed slight agreement with the F-scale items.

Further results were obtained on eight additional samples. All correlations were positive although there was great variability. A method for determining the significance of response set was developed and data from an additional representative nation-wide sample of adults with a college background indicated no significant degree of acquiescence. This was also true of our most sophisticated samples but appreciable amounts of acquiescence were found in an undergraduate sample. Possible implications of this finding for research using the F scale on undergraduate samples were noted. It is concluded that the data do not support the identification of acquiescence with authoritarianism.

## REFERENCES


1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The Authoritarian Personality*, New York: Harper, 1950.
2. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616-623.
3. BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296-299.
4. BASS, B. M. Reply to Messick and Jackson's comments on authoritarianism or acquiescence. *J. abnorm. soc. Psychol.*, 1957, **54**, 426-427.
5. CAREY, GLORIA, ROGOW, A. A., & FARRELL, CALISTA. The relationship between the F scale and aphorism usage and agreement. *J. Psychol.*, 1957, **43**, 163-167.
6. CAMPBELL, D. T., & MCCANDLESS, B. R. Ethnocentrism, xenophobia, and personality. *Human Relat.*, 1951, **4**, 185-192.
7. CHAPMAN, L. J., & CAMPBELL, D. T. Response set in the F scale. *J. abnorm. soc. Psychol.*, 1957, **54**, 129-132.
8. CHRISTIE, R., & GARCIA, J. Subcultural variation in authoritarian personality. *J. abnorm. soc. Psychol.*, 1951, **46**, 457-469.
9. COHN, T. S. The relation of the F scale to a response set to answer positively. *Amer. Psychologist*, 1953, **8**, 335. (Abstract)
10. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475-494.
11. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3-31.
12. GAGE, N. L., LEAVITT, G. S., & STONE, G. C. The psychological meaning of acquiescent set for authoritarianism. *J. abnorm. soc. Psychol.* 1957, **55**, 98-103.
13. JACKSON, D. N., & MESSICK, S. J. A note on 'ethnocentrism' and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54**, 132-134.
14. JACKSON, D. N., MESSICK, S. J., & SOLLEY, C. M. How 'rigid' is the 'authoritarian'? *J. abnorm. soc. Psychol.*, 1957, **54**, 137-140.
15. JONES, M. B. Authoritarianism and intolerance of fluctuation. *J. abnorm. soc. Psychol.*, 1955, **50**, 125-126.
16. JACOB, P. E. *Changing values in college*. New Haven: Hazen Foundation, 1956.
17. LEAVITT, H. J., HAX, H., & ROCHE, J. H. Authoritarianism and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215-221.
18. LORGE, I. Gen-like: halo or reality. *Psychol. Bull.*, 1937, **34**, 545-546. (Abstract)
19. MEER, S. J. Authoritarian attitudes and dreams. *J. abnorm. soc. Psychol.*, 1955, **51**, 74-78.
20. MESSICK, S. J., & JACKSON, D. N. Authoritarianism or acquiescence in Bass's data. *J. abnorm. soc. Psychol.*, 1957, **54**, 424-425.
21. NEWCOMB, T. M. *Personality and social change*. New York: Dryden, 1943.
22. PERLMUTTER, H. J. Some characteristics of the xenophilic personality. *J. Psychol.*, 1954, **38**, 291-300.
23. PRENTICE, N. M. The comparability of positive and negative items in scales of ethnic prejudice. *J. abnorm. soc. Psychol.*, 1956, **52**, 420-421.
24. PROTHRO, E. T., & MELIKIAN, L. H. The California public opinion scales in an authoritarian culture. *Publ. Opin. Quart.*, 1953, **17**, 353-362.
25. SHELLEY, H. P. Response set and the California attitude scales. *Educ. psychol. Measmt*, 1956, **16**, 63-67.
26. WEBSTER, H. Some quantitative results. *J. soc. Issues*, 1956, **12**, (4), 29-43.


Received March 7, 1957.

# DIFFERENTIAL AUTONOMIC RESPONSIVENESS IN THE PROCESS-REACTIVE CLASSIFICATION OF SCHIZOPHRENIA[1]

GERALD F. KING

*Michigan State University*

CONSIDERABLE research attention has recently been focused on the relationship between autonomic nervous system (ANS) activity and schizophrenic behavior. Funkenstein and collaborators (9, 10, 11, 12) have been able to predict prognosis to electric shock and other forms of therapy with a variety of neuropsychiatric patients on the basis of ANS reactivity to certain drugs. Confirmatory evidence for the relationship between ANS activity and prognosis in schizophrenia has been provided in studies by Geocaris and Kooiker (14) and Hirschstein (15). In the initial method of analyzing ANS activity (9), changes in the systolic blood pressure after adrenergic stimulation (intravenous epinephrine) and cholinergic stimulation (intramuscular mecholyl) were classified into seven patterns of response. The number of blood pressure patterns has been reduced from seven to three in subsequent research (23). Further modifications have been introduced in regard to mecholyl; response to this drug has been expressed in quantitative terms, with measures of the extent and duration of blood pressure changes being employed (13, 24).

In attempting an integration of physiological and psychological processes, Meadow and Funkenstein (23) found a relationship between the ANS response patterns of schizophrenics and impairment in abstract thinking. Such a relationship was predictable on the basis of earlier studies by Bolles *et al.* (4) and others, who concluded that poor prognosis was associated with low performance on tests of abstract thinking. An investigation by Ficca (8) revealed that neuropsychiatric patients with "unfavorable" ANS patterns of response, in terms of prognosis, produced significantly more "schizophrenic" signs on the Rorschach than patients with "favorable" ANS responses. In a later study by Meadow and his associates (24), in which only mecholyl stimulation was used, measurements of ANS activity were obtained on what might be called a "level of autonomic responsiveness" dimension (e.g., maximum fall in systolic blood pressure). A significant positive relationship was found between level of autonomic responsiveness and level of abstract thinking with schizophrenic subjects. The authors interpret the results as suggesting two polar types of schizophrenia, with the descriptions corresponding closely to the so-called "process" and "reactive" syndromes. The purpose of the present research was to explore further the relationship between level of autonomic responsiveness and the process-reactive classification.

The following gives a brief delineation of the process-reactive concept, which is usually viewed as a dichotomy: process schizophrenia is indicated by a gradual onset of psychosis and a relatively inadequate prepsychotic personality, whereas reactive schizophrenia is characterized by sudden onset of illness in a relatively normal prepsychotic personality. Additional criteria are given by Kantor, Wallner, and Winder (18). The distinction, for a particular patient, is commonly made on the basis of personal history information.

A number of investigators (17, 20, 25, 27) have presented evidence showing that schizophrenics with process characteristics have lower rates of recovery than patients more reactive in nature. In a study by Kantor, Wallner, and Winder (18), the Rorschach protocols of process schizophrenics were judged as being more "psychotic" than those of reactive schizophrenics. Brackbill and Fine (6) found that process schizophrenics could not be differentiated from organics on the basis of Rorschach indices of organicity, but that reactive schizophrenics showed significantly fewer signs of organic involvement than both groups. Becker (3), conceiving of a process-reactive *continuum*, found a significant relationship between this continuum and "genetic" scores derived from the Rorschach. The data also revealed a relationship between the

[1] This research was initiated while the author was on the staff of Veterans Administration Hospital, Battle Creek, Michigan. Grateful acknowledgment is due David W. Merrell for his early collaboration, Harry VanderKamp for providing most of the physiological measures, and Frank Restle for his interest and suggestions.

process-reactive dimension and the conceptual level of the responses to a proverbs test, but the relationship was restricted to male patients, not appearing in females. It can be summarized that, in comparison with process schizophrenics, reactives (*a*) have a more favorable prognosis, (*b*) produce Rorschach records which are less apt to be judged as "psychotic," (*c*) respond with fewer organic signs on the Rorschach, (*d*) give responses of a higher genetic level on the Rorschach, and (*e*) possibly interpret proverbs on a higher conceptual level.

The available physiological and psychological evidence suggests that the process-reactive concept can be used as a frame of reference for predicting ANS responsiveness. The following hypothesis is offered: Predominantly reactive schizophrenics exhibit a higher level of autonomic responsiveness after the injection of mecholyl than predominantly process schizophrenics. The modifying adverb "predominantly" makes the hypothesis more congruent with Becker's method of analysis (3). The variety of criteria used in making the process-reactive distinction permits patients to display a mixture of both characteristics. What is involved is a typology of personal history incidents but not a typology of patients. Thus, a given patient may fall anywhere on a scale from "purely process" to "purely reactive." To be presented in this paper are two studies which bear on the hypothesis indicated.

## STUDY I

### *Subjects*

The basic group of *Ss* consisted of 136 recently hospitalized neuropsychiatric patients, all being given the diagnosis of schizophrenia. A variety of diagnostic subtypes and symptoms were represented. None of the patients showed any evidence of neurological complications from routine examination. Ages ranged from 20 to 42 years. All patients had been hospitalized eight weeks or less when the physiological measures were obtained.

### *Physiological Measures*

The method of securing the physiological measures of ANS activity was a modification of that used by Funkenstein, Greenblatt, and Solomon (9). While the patient was lying in bed shortly after awaking in the morning, the resting systolic blood pressure was determined by a cuff manometer. The patient then received 10 mg. of mecholyl intramuscularly, and the systolic blood pressure was recorded at intervals of 2, 5, 7, 10, 12, 15, and 20 minutes. The following two measures of level of autonomic responsiveness, as

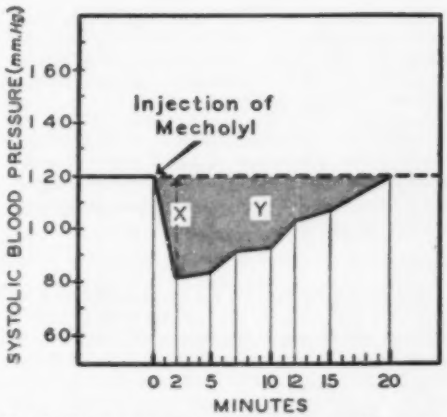


FIG. 1. DIAGRAMMATIC ILLUSTRATION OF RESPONSE OF SYSTOLIC BLOOD PRESSURE TO MECHOLYL (ADAPTED FROM MEADOW *et al.* [24])

Maximum fall in blood pressure is indicated by X. Area under baseline of resting blood pressure is indicated by Y.

derived from Meadow *et al.* (24), were obtained: (*a*) maximum fall in systolic blood pressure below the resting blood pressure following the injection of mecholyl and (*b*) area under the base-line (projection of the resting blood pressure) demarcated by blood pressure readings following mecholyl. A graphic illustration of these measures is given in Fig. 1.

The correlation between the two measures of autonomic responsiveness was found to be .82 ($N = 28$), indicating a high degree of commonality. Of the two measures, maximum fall in systolic blood pressure (MFBP) was much easier to tabulate, involving merely the subtraction of the lowest systolic blood pressure reading from the basal reading. In the present research, only MFBP was employed as a measure of level of autonomic responsiveness. Indirect evidence of the reliability of this measure is indicated by its being positively correlated with level of abstract thinking in two independent samples (24).

### *Procedure*

A sample of 60 was drawn from the basic group of schizophrenic *Ss*. All *Ss* were classified on the basis of case histories as either "process" or "reactive" by the present investigator, using the frame of reference provided by Kantor, Wallner, and Winder (18). Independent classification of all *Ss* was accomplished by either one of two other raters. None of the raters had knowledge of the physiological measures. The percentage of agreement between the two judgments was 76.7. Analysis of the agreement between the judgments with a fourfold table yielded a chi square of 17.47, which is significant at the .001 level of confidence. This is essentially of the same order of interrater agreement reported previously (18). Discussion among the raters revealed that at least two factors probably lowered reliability: (*a*) instances of inadequate case histories and (*b*) case histories that indicated process and reactive characteristics to an almost equal degree.

Using only *Ss* on which there was classificatory agreement resulted in process and reactive groups consisting of 22 and 24 *Ss*, respectively. A comparison of the groups revealed no significant difference in age. Further, in considering the process-reactive concept as a continuum, 16 *Ss* were randomly selected from the two groups and ranked accordingly by two independent raters. Rank 1 was assigned to the *S* judged to be the most process in nature, with rank 16 being occupied by the *S* being most clearly reactive. Agreement between the raters is indicated by a rho of .89.

### *Results*

The mean MFBP score of the reactive schizophrenics was 24.41 ($SD$ = 8.25), as compared to 18.50 ($SD$ = 7.71) for the process schizophrenics.[9] The usual $t$ test gives a value of 2.51, which indicates that this difference is significant at the .01 level (one-tailed test). For the 16 *Ss*, the rho between the sets of ranks on the process-reactive dimension and MFBP is −.58 ($p <$ .01, one-tailed test). In accordance with the hypothesis, the reactive schizophrenics displayed a higher level of autonomic responsiveness following the injection of mecholyl than the process schizophrenics. The results also indicate that it is feasible to order the process and reactive syndromes on a continuum.

## STUDY II

### *Subjects*

With the elimination of the 46 *Ss* who participated in Study I, the basic group of *Ss* was reduced to 90.

### *Physiological Measures*

The same method as in Study I was employed.

### *Procedure*

The following modification was introduced in the classificatory procedure: all 90 *Ss* were classified as either "process," "process-reactive," or "reactive." The process-reactive category was designed to be an intermediate one, representing schizophrenics with both types of characteristics to an almost equal degree. Otherwise, the same rating procedure was followed as in Study I. The percentage of agreement between the independent judgments was 87.8. The corresponding chi square, using a 3-by-3 contingency table, was 121.53, which is significant considerably beyond the .001 level. Groups representing the three categories were formed from the *Ss* on which there was classificatory agreement. Balanced for $N$, each group consisted of 24 *Ss*. A comparison for age yielded no significant differences among the groups.

In addition, scores on the Elgin Prognostic Scale (31) were obtained for 22 *Ss*. This scale consists of ratings on factors which are based on the process-reactive differentiation. Data for the ratings were secured from interviews and clinical folders (case histories, etc.). Higher scores on the scale represent a greater incidence of process elements. Wittman (31) reports an interrater reliability coefficient of .87 for an $N$ of 61. In the present research, 12 *Ss* were rated independently by two raters, the interrater rho being .94.

### *Results*

Treating the MFBP scores of the three groups by analysis of variance yielded an $F$ ratio that is significant beyond the .01 level. The mean MFBP scores were of ascending magnitude, with the process schizophrenics being lowest ($M$ = 17.04), the process-reactives intermediate ($M$ = 22.79), and the reactives highest ($M$ = 26.62). This trend is in accordance with the hypothesis and the conception of a process-reactive continuum. Since this is not a *post hoc* analysis, there is justification for computing $t$ tests between the individual means. Using one-tailed tests, the results show that the process schizophrenics were significantly lower than the process-reactives ($t$ = 2.39, $p < .025$), who in turn approached being significantly lower than the reactive schizophrenics ($t$ = 1.54, $p < .07$). The rho between the Elgin scale and the MFBP scores ($N$ = 22) was −.49 ($p < .025$, one-tailed test). The findings thus indicate that the reactive schizophrenics exhibited a greater autonomic response to mecholyl than the process schizophrenics; or, in other words, the process-reactive dimension is inversely related to level of autonomic responsiveness. As in Study I, the hypothesis is supported.

## DISCUSSION

The results provide a basis for discussing and speculating about some of the past research dealing with physiological aspects of schizophrenia. Citing a variety of supporting studies, Angyal, Freeman, and Hoskins (2) and Hoskins (16) characterize schizophrenics, in comparison with normals, as showing reduced physiological responsiveness. However, confirmation of this position has not been obtained in a number of other studies (1, 7, 26). As suggested by Sands and Rodnick (29) and as illustrated by Malmo, Shagass, and Smith (22), contradictory results in this area are undoubtedly due in part to variations in the types of experimental stimulation and the nature of the physiological measures. Variations in the schizophrenic samples employed in different studies is another likely factor. According to the present research, responsiveness is related to the process-reactive make-up of the schizophrenic sample. If we consider pure cases of "process" and "reactive" patients for the sake of discussion, the evidence points to diminished physiological responsiveness in process but not in reactive schizophrenia. Thus, reduced responsiveness would not be found in studies where the schizophrenic subjects were predominantly reactive in character.

[9] In comparison with the distribution reported by Funkenstein *et al.* (13), the MFBP scores obtained in the present research seem to be depressed.

One possible reaction to the preceding discussion is that it moves in the direction of a two-factor theory of schizophrenia. Process characteristics are associated with a physiological "impairment," at least implicitly. What remains, it would seem, is to tie the disturbance in the reactive syndrome to interpersonal factors. This notion is compatible with the contention of Brackbill (5), who, after reviewing the studies of brain dysfunction in schizophrenia, postulates the existence of central nervous system pathology in process but not in reactive patients. In this context, the question arises as to whether autonomic disturbances are primary or secondary to cortical defect.

If the information on normal responsiveness is given more consideration, it is possible to cast the problem in a different light. Although the data on the reactivity of normals to mecholyl are meagre (10, 14), it seems to indicate that their blood pressure responses are very similar to what can be expected from process schizophrenics. In graphic terms, if level of autonomic responsiveness represents the ordinate and normals, reactives, and process schizophrenics lie on the abscissa in that order, the relationship can be described as curvilinear. In accordance with this hypothetical graph, the reactives would be high in autonomic responsiveness, whereas the normals and process patients would be relatively low. Such a relationship offers a more adequate account of studies which have found that schizophrenics, especially patients in the early stages of illness, are physiologically even more reactive than normals (e.g., 21, 30). A study by Pfister (28) gives a more complete illustration of the curvilinear relationship: Early schizophrenics were hyper-reactive on cardiovascular measures; if the illness persisted, however, the same *S*s dropped toward or even below the normal level. Although a loose interchange of concepts is sometimes necessary, a good deal of the available data can be fitted to the curve. A possible explanation lies in viewing both normals and process schizophrenics as representing relatively uniform, stable adjustments over a period of time. It is suggested that the corresponding physiological homeostasis is indicated by a low autonomic responsiveness. A heightened autonomic reactivity then would reflect a disruption of homeostasis, which corresponds with the more transient, precipitous nature of the reactive adjustment.

Although problems of interpretation are posed, the results in general provide additional evidence of the meaningful nature of the process-reactive classification, or dimension, of schizophrenia. Using this concept of schizophrenia, which cuts across symptomatology, evidence has been compiled from this and other studies revealing interrelationships among several classes of variables: case history information, psychological dimensions, physiological measures, and prognosis. The numerous failures of past studies dealing with schizophrenia seem to give this classification even more significance. When samples of undifferentiated schizophrenics, or even ones based on the standard diagnostic subtypes, have been employed in research, the findings have typically been either negative, not readily interpretable, or not reproducible (19). Pending the development of a better schema for delineating the heterogeneous category of schizophrenia the process-reactive classification seems to offer a productive orientation for research.

## SUMMARY

This investigation consisted of two studies which were concerned with the relationship between the process-reactive classification of schizophrenia and autonomic nervous system activity. On the basis of psychological and physiological findings, it was hypothesized that predominately reactive schizophrenics display a higher level of autonomic responsiveness after the injection of mecholyl than predominantly process schizophrenics. Similar in nature, the two studies had the following methodological features in common: (*a*) The process-reactive classification of schizophrenic *S*s was made on the basis of case history information, (*b*) a small sample of the total *N* was ranked on a process-reactive continuum, and (*c*) maximum fall in systolic blood pressure constituted the measure of autonomic responsiveness.

In accordance with the hypothesis, the results of both studies revealed that the patients classified as "reactive" exhibited a significantly greater fall in blood pressure after the administration of mecholyl than the process patients. Significant rank-order correlations were also obtained between the process-reactive dimen-

sions employed in each study and maximum fall in blood pressure, indicating that it is feasible to order the process and reactive syndromes on a continuum. Some implications of the findings in regard to physiological responsiveness in schizophrenia were discussed. It was concluded that the process-reactive classification seems to provide a fruitful frame of reference for research in schizophrenia.

## REFERENCES


1. ALTSCHULE, M. D., & SULZBACH, W. M. Effect of carbon dioxide on acrocyanosis in schizophrenia. *Arch. Neurol. Psychiat.*, 1949, **61**, 44–55.
2. ANGYAL, A., FREEMAN, H., & HOSKINS, R. G. Physiologic aspects of schizophrenic withdrawal. *Arch. Neurol. Psychiat.*, 1940, **44**, 621–626.
3. BECKER, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **53**, 229–236.
4. BOLLES, M. M., ROSEN, G. P., & LANDIS, C. Psychological performance tests as prognostic agents for the efficacy of insulin therapy in schizophrenia. *Psychiat. Quart.*, 1938, **12**, 733–737.
5. BRACKBILL, G. A. Studies of brain dysfunction in schizophrenia. *Psychol. Bull.*, 1956, **53**, 210–226.
6. BRACKBILL, G. A., & FINE, H. J. Schizophrenia and central nervous system pathology. *J. abnorm. soc. Psychol.*, 1956, **52**, 310–313.
7. CAMERON, D. E. Heat production and heat control in the schizophrenic reaction. *Arch. Neurol. Psychiat.*, 1934, **32**, 704–711.
8. FICCA, S. C. Relationship of "autonomic" blood pressure pattern types and subjects' performance on the Wechsler-Bellevue and the Rorschach test. Unpublished doctoral dissertation, Pennsylvania State Univer., 1950.
9. FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. Autonomic nervous system changes following electric shock treatment. *J. nerv. ment. Dis.*, 1948, **108**, 409–422.
10. FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. Psycho-physiological study of mentally ill patients. *Amer. J. Psychiat.*, 1949, **106**, 16–28.
11. FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. Autonomic changes paralleling psychologic changes in mentally ill patients. *J. nerv. ment. Dis.*, 1951, **114**, 1–18.
12. FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. An autonomic nervous system test of prognostic significance in relation to electroshock treatment. *Psychosom. Med.*, 1952, **14**, 347–362.
13. FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. Nor-epinephrine-like and epinephrine-like substances in psychotic and psychoneurotic patients. *Amer. J. Psychiat.*, 1952, **108**, 652–662.
14. GEOCARIS, K. H., & KOOIKER, J. E. Blood pressure responses of chronic schizophrenic patients to epinephrine and mecholyl. *Amer. J. Psychiat.*, 1956, **112**, 808–813.
15. HIRSCHSTEIN, R. The significance of characteristic autonomic nervous system responses in the adjustment, change and outcome in schizophrenia. *J. nerv. ment. Dis.*, 1955, **122**, 254–262.
16. HOSKINS, R. G. *The biology of schizophrenia*. New York: Norton, 1946.
17. KANT, O. Differential diagnosis of schizophrenia in light of concepts of personality-stratification. *Amer. J. Psychiat.*, 1940, **97**, 342–357.
18. KANTOR, R. E., WALLNER, J. M., & WINDER, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17**, 157–162.
19. KING, G. F. Research with neuropsychiatric samples. *J. Psychol.*, 1954, **38**, 383–387.
20. LANGFELDT, G. The diagnosis of schizophrenia. *Amer. J. Psychiat.*, 1951, **108**, 123–125.
21. MALMO, R. B., & SHAGASS, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, **11**, 9–24.
22. MALMO, R. B., SHAGASS, C., & SMITH, A. A. Responsiveness in chronic schizophrenia. *J. Pers.* 1951, **19**, 359–375.
23. MEADOW, A., & FUNKENSTEIN, D. H. The relationship of abstract thinking to the autonomic nervous system in schizophrenia. In P. Hoch & J. Zubin (Eds.), *Relation of psychological tests to psychiatry*. New York: Grune & Stratton, 1952. Pp. 131–149.
24. MEADOW, A., GREENBLATT, M., FUNKENSTEIN, D. H., & SOLOMON, H. C. Relationship between capacity for abstraction in schizophrenia and physiologic response to autonomic drugs. *J. nerv. ment. Dis.*, 1953, **118**, 332–338.
25. MILICI, P. Postemotive schizophrenia. *Psychiat. Quart.* 1939, **13**, 278–293.
26. PARSONS, E. H., GILDEA, E. F., RONZONI, E., & HULBERT, S. Z. Comparative lymphocytic and biochemical responses of patients with schizophrenia and affective disorders to electroshock, insulin shock, and epinephrine. *Amer. J. Psychiat.*, 1949, **105**, 573–580.
27. PASKIND, H. A., & BROWN, M. Psychoses resembling schizophrenia occurring with emotional stress and ending in recovery. *Amer. J. Psychiat.*, 1940, **96**, 1379–1388.
28. PFISTER, H. O. Disturbances of the autonomic nervous system in schizophrenia and their relations to the insulin, cardiazol and sleep treatments. *Amer. J. Psychiat. Suppl.*, 1938, **109**, 94–118.
29. SANDS, S. L., & RODNICK, E. H. Concept and experimental design in the study of stress and personality. *Amer. J. Psychiat.*, 1950, **106**, 673–679.
30. WILLIAMS, M. Psychophysiological responsiveness to psychological stress in early chronic schizophrenic reactions. *Psychosom. Med.*, 1953, **15**, 456–462.
31. WITTMAN, PHYLLIS. Scale for measuring prognosis in schizophrenic patients. *Elgin State Hosp. Papers*, 1941, **4**, 20–33.


Received May 31, 1956.

# AMBIGUITY AND ANXIETY[1]

ANDREW S. DIBNER
*Clark University*

THE concept of anxiety has been subject to considerable scrutiny in the last few years. The present study is concerned with the meaning of anxiety, where anxiety is considered as a momentary state, not as a "personal constant" (14) or as a disease process (17). The thesis presented is that anxiety is directly related to the degree of ambiguity in the situation to which the individual must make some adjustive reaction, that it is the condition of *uncertainty* (of perception, expectation, success, prediction, control, or even survival) which is the essence of anxiety, and that the probability of uncertainty is greater whenever external conditions are ambiguous. The ambiguous field permits a variety of possible structures (20), offering the individual a complex array of choices with a lessened probability of correctly predicting the results of his decisions because of a lack of definitive cues. Cues may be absent, vague, and inconsistent or contradictory. A highly structured field, on the other hand, provides guide lines; the individual may properly evaluate it and behave toward it in a meaningful way. There is no implication that ambiguity is always a sufficient condition for anxiety; the present argument bears only on situations containing some element of personal importance or ego involvement and calling for an adjustive response.

In this paper the terms "ambiguity" and "structure" are used to describe both objective properties of the stimulus field and psychological interpretations of the field. The stimulus field is said to have greater structure according to the extent to which it contains sufficient perceivable characteristics to be subject to a greater consensus of interpretations by observers. It is relatively ambiguous (or unstructured) if its characteristics are less sharply defined, permitting more alternative interpretations. Thus, one can diminish the structure of a stimulus field in an objective way by subtracting information from it, or one can add to structure (diminish ambiguity) by adding information. On the other hand, the psychological interpretation of the structure of a stimulus field (Subjective Structure as opposed to Objective Structure) has as its determinants not only the structural properties of the field but also those intrapersonal attributes which affect perception. Persons differ in their tendency to impose structure on an ambiguous field (1, 15, 24), and it seems probable that even a rather highly structured stimulus may be perceived as ambiguous by some.

There have been few experiments related to such hypotheses about the relationship of anxiety to objective and subjective structure (7, 13, 28) and none in a natural field situation. This study utilized five measures of anxiety to test patients' reactions to psychological interviews which were varied in the degree of structuring provided by the interviewers.

## METHOD

### *Subjects*

The 40 *S*s were male veterans selected from consecutive admissions to the Neuropsychiatric Service of the Dearborn Veterans Administration Hospital.[2] All cases of suspected psychosis or organic brain disease were excluded, as were those involving poor verbal ability or low IQ. Selected *S*s were all cases of neurosis, psychosomatic disease, or psychopathic personality. They had to be ambulatory and in sufficient control to be on the open wards, relatively new arrivals in the hospital, no more than 55 years of age, and without previous contacts with the Clinical Psychology section before the experimental interview. All were seen within one week of admission and were randomly assigned to interviewer and conditions. A later check revealed equivalence of the groups of *S*s receiving Structured or Ambiguous interviews on such variables as age (Range 25–54, Mean 31.7), diagnosis, race, education (Mean 10.5 years), IQ (Mean 105.4), marital status, number of days in the hospital before being interviewed, sedation on the day of the interview, time of day of inter-

[1] This paper is based upon a doctoral dissertation done at the University of Michigan (9). The writer wishes to express his appreciation to Edward S. Bordin for his stimulating guidance. Some support was received from the Research Project: Analysis of Therapeutic Interaction, USPHS M-516, National Institute of Mental Health, Bethesda, Maryland.

[2] The writer is indebted to Thomas P. Crane, Manager, and J. J. Lasky, then Chief Clinical Psychologist, for their help.

view, and the number and type of previous psychiatric or psychological contacts.

*Procedure*

Four psychologists[2] each conducted clinical evaluative interviews, five in an ambiguous manner and five in a structured manner, thus creating a group of 20 Ambiguous (A) and 20 Structured (S) interviews. The interviews were tape recorded and readings were made of the patients' skin resistance throughout the interviews. Following the interviews, the patients were asked standardized questions by *E*, and completed check lists and cardsorts to obtain their reactions to the sessions. Five anxiety measures were compared with an objective index of structuredness of interviewer behavior and with a measure of patients' perception of structure.

*Conditions.* Situational factors of the interview setting were strictly controlled. *S*s were not told that the interviews were experimental; the recording and skin resistance apparatus were explained as standard hospital techniques. The post-interview data collecting, however, was justified to the *S*s as a research study. The interviewers were trained in the two conditions. In the A condition, the patient was told it was considered best if he talked about himself in any way he wished, saying whatever he thought would be of help to the interviewer in understanding him and his life. The interviewer accepted anything brought up and supported when necessary, but tried to avoid giving any but the most general cues as to what to talk about. He tried to avoid giving advice or asking follow-up questions. In the S condition, the interviewer asked questions in predetermined areas and followed up any lead the patient offered. He was free to make his observations or feelings about the material known to the patient in the form of questions, information, or counsel whenever he wished.

*Measures of Structure*

*Objective Structure.* The degree of structuring by the interviewers was measured by applying Osburn's method (23) to 10-minute samples of the typescripts. This index of degree of structure in an interview is based on a weighted coding of individual interviewer responses and on the total number of his responses. The coding considers interviewer responses in terms of their direction (information-giving or information-seeking), referent of the response (client referent or situation referent) and generality of the response (most general to highly specific in content). Since the distribution of these scores was considerably skewed, it was necessary to use nonparametric statistics in the analysis of results.

*Subjective Structure.* Subjective structure was defined by clinical judges' ratings of the degree of ambiguity as seen by the patient. Ratings were based on analysis of transcripts of patients' responses to three open-ended questions following their interviews. The questions asked how the patient felt the interview went, how it could have gone better, and if he felt he knew what to talk about in the interview. Instructions to raters follow:

> Following a clinical first interview with a psychologist, patients were asked about it by the researcher. Certain questions asked them were designed to elicit *their perception* of the degree of structuredness or ambiguity of the interview situation they had just faced. Degree of structuredness of the interview *as experienced by the patient* is what is being considered now.
>
> The patients had been faced with the task of telling the psychologist about themselves. What they should or should not tell the psychologist is determined by many factors. For example, the degree of defining or structuring by the interviewer is a factor. Another factor involves the ideas the patient has about what is important in his case and what he thinks the interviewer is interested in. Also the *needs* the patient might have to reveal some things and keep others secret would affect his perception of what should be talked about. Thus, with some factors coming from within the patient, his *perception* of the structure of the task might be different from the perception other patients might have facing the same interviewer in the same situation. Even though an interviewer may feel he is providing many cues as to what the patient should talk about, some patients might still feel as if the situation is very ambiguous. Others might see their interview as cut-and-dried, highly structured, even though their interviewer doesn't ask them one specific question. They might have a structure of their own that they apply.
>
> The responses to certain questions by nine patients are presented here. Please rate for each patient how you think he perceived the *structuredness* of the interview he had. Rate each interview by placing check marks on the 6-point linear scales on the rating sheet. You can check between numbers if unsure of the placement.

[2] The author is grateful to Richard Cutler, Barbara Nachmann, Stanley Segal, and David Shapiro for their cooperation as interviewers.

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| Ambiguous. The patient felt he didn't know what to talk about in the interview he had. | | | | | Structured. He had no doubts at all concerning what he should talk about. |

Rank intercorrelation of five judges rating nine patients was .56. *E* rated all 40 patients. His rank correlation with the median of the five raters was .90.

*Measures of Anxiety*

*Palmar Skin Conductance.* Resistance readings were recorded at 30-second intervals by the *E* in a room adjoining the interview room. Following an accommodation period of four minutes, a curve of conductance was plotted for each patient over a 35-minute period. The slope of the curve from the mean of the first 2½ minutes to the mean of the final 10 minutes was used as the skin conductance measure of anxiety for the interview. The mean and range of conductance was equivalent to data reported by others (28, 31). Range of conductance-change through the interviews was from −180 to 952 micromhos, with mean change of 148.5.

*Anxiety Cards.* Following the interview, *E* presented each *S* with 20 cards, one at a time, each containing phrases describing feelings of tension or absence of tension or common physiological concomitants of tension. The *S* was asked to place each card on one of three piles indicating that he "Felt that way," "Was not sure," or "Did not feel that way." Statements on the Anxiety Cards appear below. The starred items are nonanxious in character.

1. I felt I couldn't concentrate too well.
2. I felt I was sweating quite a bit.
3. During most of the talk I felt perfectly at ease.*
4. I felt restless most of the time.
5. I felt my stomach was "turning over".
6. I kept wondering what this interview was all about.
7. I felt clear-headed all the way through the talk.*
8. After about five minutes I felt relaxed and felt relaxed till the talk was over.*
9. I felt nervous through the whole talk.
10. The talk left me feeling tired.
11. I felt too hot.
12. I was worried about what I was telling him.
13. I felt pains in my chest sometimes while we were talking.
14. I had difficulty in breathing sometimes while we were talking.
15. I felt very comfortable from the beginning of the talk right through to the end.*
16. Sometimes during the talk I could hear my heart pounding.
17. Sometimes my throat felt dry while we were talking.
18. I felt my muscles tighten up sometimes during the talk.
19. I felt much better during this talk than when I talk to most doctors.*
20. Most of the time I felt rather tense.

These statements were selected on the basis of a pilot study using a longer list of anxiety statements. The five Nonanxious statements were included with the 15 Anxious statements to permit *S* to profess control of anxiety, but only the number of cards bearing anxious statements that the patient sorted as "Felt that way" yielded his Anxiety Card score. The split-half reliability of the scale was .76. An item analysis of the anxious cards further showed that the scale had high internal consistency. Obtained scores ranged from 0 to 11 with a mean of 3.15.

*Anxiety Ratings.* Ratings of anxiety were made by clinical judges from transcripts of 10-minute samples of the interviews. Four VA staff psychologists were given 17 10-minute samples of transcribed interviews and were asked to rate each patient response or 10-line segment of long responses on a 5-point scale of anxiety. The sample of 17 cases included two for each interviewer for each condition, plus one randomly selected case. Instructions to raters appear below:

Read through the transcript and rate each patient response (or 10-line segment of long responses as indicated). Use the following 5-point scale of Anxiety:

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Relaxed, not anxious | | Anxious | | Very anxious |

Try to consider the scale as linear and continuous with 3 as the center-point representing the average anxiety level common to patients having their first clinical interview. You may check between the numbers if unsure of placement of some ratings.

In reading each patient statement try to get a feel for his emotional state at the time he is saying those words. Content may or may not be important. Try to judge mainly by the way he verbalizes the content. Here are some suggested cues. Also use that "3rd ear" freely to pick up the patient's emotional tone.

Some signs of anxiety:

Fumbling around with how to present thoughts.
Sighing, tense laughter, changes in pitch or volume.
Tripping over words, stuttering, leaving words or thoughts half finished.
Patient stating he is uncomfortable or apprehensive.
Angry tone that sounds like reacting to anxiety.

Some signs of little anxiety:

view, and the number and type of previous psychiatric or psychological contacts.

*Procedure*

Four psychologists[3] each conducted clinical evaluative interviews, five in an ambiguous manner and five in a structured manner, thus creating a group of 20 Ambiguous (A) and 20 Structured (S) interviews. The interviews were tape recorded and readings were made of the patients' skin resistance throughout the interviews. Following the interviews, the patients were asked standardized questions by *E*, and completed check lists and cardsorts to obtain their reactions to the sessions. Five anxiety measures were compared with an objective index of structuredness of interviewer behavior and with a measure of patients' perception of structure.

*Conditions.* Situational factors of the interview setting were strictly controlled. *S*s were not told that the interviews were experimental; the recording and skin resistance apparatus were explained as standard hospital techniques. The post-interview data collecting, however, was justified to the *S*s as a research study. The interviewers were trained in the two conditions. In the A condition, the patient was told it was considered best if he talked about himself in any way he wished, saying whatever he thought would be of help to the interviewer in understanding him and his life. The interviewer accepted anything brought up and supported when necessary, but tried to avoid giving any but the most general cues as to what to talk about. He tried to avoid giving advice or asking follow-up questions. In the S condition, the interviewer asked questions in predetermined areas and followed up any lead the patient offered. He was free to make his observations or feelings about the material known to the patient in the form of questions, information, or counsel whenever he wished.

*Measures of Structure*

*Objective Structure.* The degree of structuring by the interviewers was measured by applying Osburn's method (23) to 10-minute samples of the typescripts. This index of degree of structure in an interview is based on a weighted coding of individual interviewer responses and on the total number of his responses. The coding considers interviewer responses in terms of their direction (information-giving or information-seeking), referent of the response (client referent or situation referent) and generality of the response (most general to highly specific in content). Since the distribution of these scores was considerably skewed, it was necessary to use nonparametric statistics in the analysis of results.

*Subjective Structure.* Subjective structure was defined by clinical judges' ratings of the degree of ambiguity as seen by the patient. Ratings were based on analysis of transcripts of patients' responses to three open-ended questions following their interviews. The questions asked how the patient felt the interview went, how it could have gone better, and if he felt he knew what to talk about in the interview. Instructions to raters follow:

> Following a clinical first interview with a psychologist, patients were asked about it by the researcher. Certain questions asked them were designed to elicit *their perception* of the degree of structuredness or ambiguity of the interview situation they had just faced. Degree of structuredness of the interview *as experienced by the patient* is what is being considered now.
>
> The patients had been faced with the task of telling the psychologist about themselves. What they should or should not tell the psychologist is determined by many factors. For example, the degree of defining or structuring by the interviewer is a factor. Another factor involves the ideas the patient has about what is important in his case and what he thinks the interviewer is interested in. Also the *needs* the patient might have to reveal some things and keep others secret would affect his perception of what should be talked about. Thus, with some factors coming from within the patient, his *perception* of the structure of the task might be different from the perception other patients might have facing the same interviewer in the same situation. Even though an interviewer may feel he is providing many cues as to what the patient should talk about, some patients might still feel as if the situation is very ambiguous. Others might see their interview as cut-and-dried, highly structured, even though their interviewer doesn't ask them one specific question. They might have a structure of their own that they apply.
>
> The responses to certain questions by nine patients are presented here. Please rate for each patient how you think he perceived the *structuredness* of the interview he had. Rate each interview by placing check marks on the 6-point linear scales on the rating sheet. You can check between numbers if unsure of the placement.

[3] The author is grateful to Richard Cutler, Barbara Nachmann, Stanley Segal, and David Shapiro for their cooperation as interviewers.

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| Ambiguous. The patient felt he didn't know what to talk about in the interview he had. | | | | | Structured. He had no doubts at all concerning what he should talk about. |

Rank intercorrelation of five judges rating nine patients was .56. *E* rated all 40 patients. His rank correlation with the median of the five raters was .90.

*Measures of Anxiety*

*Palmar Skin Conductance.* Resistance readings were recorded at 30-second intervals by the *E* in a room adjoining the interview room. Following an accommodation period of four minutes, a curve of conductance was plotted for each patient over a 35-minute period. The slope of the curve from the mean of the first 2½ minutes to the mean of the final 10 minutes was used as the skin conductance measure of anxiety for the interview. The mean and range of conductance was equivalent to data reported by others (28, 31). Range of conductance-change through the interviews was from −180 to 952 micromhos, with mean change of 148.5.

*Anxiety Cards.* Following the interview, *E* presented each *S* with 20 cards, one at a time, each containing phrases describing feelings of tension or absence of tension or common physiological concomitants of tension. The *S* was asked to place each card on one of three piles indicating that he "Felt that way," "Was not sure," or "Did not feel that way." Statements on the Anxiety Cards appear below. The starred items are nonanxious in character.

1. I felt I couldn't concentrate too well.
2. I felt I was sweating quite a bit.
3. During most of the talk I felt perfectly at ease.*
4. I felt restless most of the time.
5. I felt my stomach was "turning over".
6. I kept wondering what this interview was all about.
7. I felt clear-headed all the way through the talk.*
8. After about five minutes I felt relaxed and felt relaxed till the talk was over.*
9. I felt nervous through the whole talk.
10. The talk left me feeling tired.
11. I felt too hot.
12. I was worried about what I was telling him.
13. I felt pains in my chest sometimes while we were talking.
14. I had difficulty in breathing sometimes while we were talking.
15. I felt very comfortable from the beginning of the talk right through to the end.*
16. Sometimes during the talk I could hear my heart pounding.
17. Sometimes my throat felt dry while we were talking.
18. I felt my muscles tighten up sometimes during the talk.
19. I felt much better during this talk than when I talk to most doctors.*
20. Most of the time I felt rather tense.

These statements were selected on the basis of a pilot study using a longer list of anxiety statements. The five Nonanxious statements were included with the 15 Anxious statements to permit *S* to profess control of anxiety, but only the number of cards bearing anxious statements that the patient sorted as "Felt that way" yielded his Anxiety Card score. The split-half reliability of the scale was .76. An item analysis of the anxious cards further showed that the scale had high internal consistency. Obtained scores ranged from 0 to 11 with a mean of 3.15.

*Anxiety Ratings.* Ratings of anxiety were made by clinical judges from transcripts of 10-minute samples of the interviews. Four VA staff psychologists were given 17 10-minute samples of transcribed interviews and were asked to rate each patient response or 10-line segment of long responses on a 5-point scale of anxiety. The sample of 17 cases included two for each interviewer for each condition, plus one randomly selected case. Instructions to raters appear below:

> Read through the transcript and rate each patient response (or 10-line segment of long responses as indicated). Use the following 5-point scale of Anxiety:
>
> | 1 | 2 | 3 | 4 | 5 |
> |---|---|---|---|---|
> | Relaxed, not anxious | | Anxious | | Very anxious |
>
> Try to consider the scale as linear and continuous with 3 as the center-point representing the average anxiety level common to patients having their first clinical interview. You may check between the numbers if unsure of placement of some ratings.
>
> In reading each patient statement try to get a feel for his emotional state at the time he is saying those words. Content may or may not be important. Try to judge mainly by the way he verbalizes the content. Here are some suggested cues. Also use that "3rd ear" freely to pick up the patient's emotional tone.
>
> Some signs of anxiety:
>
> Fumbling around with how to present thoughts.
> Sighing, tense laughter, changes in pitch or volume.
> Tripping over words, stuttering, leaving words or thoughts half finished.
> Patient stating he is uncomfortable or apprehensive.
> Angry tone that sounds like reacting to anxiety.
>
> Some signs of little anxiety:

Smooth running speech, thoughts connected and uninterrupted.

Statements of confidence that don't smell of defensiveness.

Some responses are very short but try to rate them as well as you can. For example: The patient may ask the interviewer to repeat (P: What?); from the context try to get a feel for what is behind this "What?" At that point the patient may be anxious or relaxed.

For each rater, the individual ratings of responses within an interview were averaged to obtain a single score for an interview. Rank intercorrelation of the four raters was .59. One of the raters, whose correlation with the median of the other three raters was .52, judged the rest of the interview samples.

Because of the possibility of bias from the rater's knowledge of the hypotheses and from his judgment as to interviewer structuring (by reading interviewer responses), a test of presence of bias was made. Three additional raters were given 16 of the interview samples with all interviewer responses, short patient responses, and any patient references to lack of structure eliminated. The three raters of these specially prepared samples had a rank intercorrelation of .60, achieving about the same degree of agreement as full sample raters. Their median ratings of the 16 interviews correlated .64 with the median ratings of the full sample raters. If corrected for attenuation due to inaccuracy of rating, this correlation would actually approach unity. The variance unaccounted for by the correlation of .64 is almost entirely a function of the unreliability of ratings, leaving little room for effect by bias on the part of the full sample raters.

*Speech Cue Scales I and II.* Two measures of clinical signs of anxiety in speech, scored from the 10-minute samples of the transcripts, were employed. Their method of development is reported elsewhere (8). The scales are based upon a count of speech cues reflecting disorganization, emotional involvement, and defensiveness. The cues included in each scale are as follows:

*Cue Scale I*

BR —Breaking in with a new thought, generally by breaking into another sentence.

UN —Unfinished sentence.

IDK—Saying "I don't know," not in answer to a direct question, but as in resignation, disgust, or despair.

INT—Interrupted sentence. The thought is interrupted within the sentence but is eventually completed in much the same manner as it was begun.

R —Repeating words or phrases when full words or phrases are repeated close in sequence within a sentence.

ST —Stuttering when repeating parts of words or when leaving a word unfinished.

*Cue Scale II*

L —Laughing. Any kind of laugh or chuckle.

V —Voice change. Change in volume, becoming louder or lower.

SI —Sighing or deep breath.

By inspection of a table of their intercorrelations, the speech cues were grouped into two patterns defining the two scales. With the exception of IDK, the cues of Scale I are all indicators of speech disruption. Scale II is composed of cues related to nonverbal voice characteristics. The agreement of two judges rating the cues grouped in Scale I was 65 per cent, the cues grouped in Scale II, 100 per cent. Cue Scales I and II have a Kendall's rank correlation (tau) of $-.03$ ($p = .81$). Unrelated, they were used independently as anxiety measures.

## RESULTS

### *Structuredness*

Applying the Objective Structure measure to the interviews revealed that the interviewers were able to adhere to the conditions. Out of the total group of 39 interviews of which transcriptions were made, all but two of the S interviews had higher Objective Structure scores than did the A interviews, and only two A interviews overlapped the S distribution. Interviewers were found by analysis of variance by ranks (16) not to differ significantly in adhering to the conditions. Thus, interviews by all interviewers could justifiably be pooled for analysis.

Subjective Structure or the patients' perception of structuredness was found to be strongly related to the experimental conditions ($r_{bis} = .71$), but the relationship was considerably lessened when individual variation was more fully represented by using a rank correlation with Objective Structure (tau = .35). Thus, a patient's perception of structure is in part determined by the structuring behavior of the interviewer, but variance appears to be introduced by the unreliability of the measures and interpatient personality differences which affect perception.

### *Measures of Anxiety*

Intercorrelations among the five anxiety measures used as independent criteria appear

TABLE 1

CORRELATION (TAU) OF ANXIETY MEASURES (40 interviews)

| | Anxiety Cards | Skin Conductance | Anxiety Ratings | Cue Scale I |
|---|---|---|---|---|
| Skin conductance | .14 | — | — | — |
| Anxiety ratings | .00 | −.02 | — | — |
| Cue Scale I | .01 | −.08 | .32** | — |
| Cue Scale II | .07 | −.14 | .26** | −.03 |

** $p < .01$.

in Table 1. Only the correlations between the Cue scales and Anxiety ratings reach significance, and these are low. Both Cue scales and the Anxiety ratings were based on the same data—transcribed patient responses. Judges doing the anxiety ratings were instructed to be aware of speech clues in their estimate of patient anxiety. Although the two Cue scales are themselves uncorrelated, both sets of speech cues enter into a judge's estimate of anxiety when reading transcripts.

*Ambiguity and Anxiety*

A test of the hypotheses of relationship between anxiety and ambiguity was made by correlating the measures of Objective and Subjective structure with the five anxiety measures across all interviews. Table 2 presents these correlations. In anxiety *vs.* structure of the situation as defined by the interviewer (Objective Structure) all correlations are negative, and four of the five reach significance, though all are low. The largest negative relationship (−.40) is found between structuredness of interviewer behavior and clinical judgment of patient anxiety. No significant relationship was found between Objective Structure and the autonomic measure of anxiety (skin conductance).

Similarly, correlations of anxiety to patients' perception of structure (Subjective Structure) are again all negative, but only two anxiety measures, the patient's self-report (Anxiety Cards) and his change in skin conductance, show significant relationships.

Thus, the hypotheses of negative relationship between anxiety and structure are borne out by four anxiety measures in the case of Objective Structure and two in the case of Subjective Structure.

In these analyses, the total group of experimental interviews was used because the measures of Objective and Subjective Structure could be applied to all interviews, ranging them on continua. However, the total group was, from another point of view, made up of two groups toward which the interviewers were instructed to behave differently, as to both their early orientation of the patient and their structuring behavior throughout the interview. It was thought that a more rigorous test of the hypotheses could be attempted by examining their utility *within* the separate A and S groups. Such an analysis would eliminate effects due to the difference in initial orientation of patients and the difference in expectations of the interviewers. Presumably, within each group (A or S) each interviewer tried to be *the same* as he was in other interviews in that group, not *different*, as he had to be when he interviewed the next patient under the other condition. Thus, a test of the hypotheses within probably more homogeneous groups would be a purer and stricter test. Because $N$ was reduced to half, providing a small sample of only 20 cases (in some cases 19), and because the range of structuredness is reduced, it would be expected that error factors would have a larger effect in such an analysis, and higher correlations would be needed to reach significance. By comparing results from the analysis of these subgroups with findings across all interviews, however, one would be perhaps better able to understand the nature of variation in the interview situation.

Table 3 presents correlations of Objective and Subjective Structure with anxiety within the experimental A and S groups. All the relationships are negative within the groups, except those between Objective Structure and Skin Conductance and between Subjective

TABLE 2

CORRELATION (TAU) OF STRUCTURE MEASURES AND ANXIETY MEASURES ACROSS ALL INTERVIEWS

| | Anxiety Ratings | Anxiety Cards | Skin Conductance | Cue Scale I | Cue Scale II |
|---|---|---|---|---|---|
| Objective structure | −.40** | −.19* | −.05 | −.18* | −.23* |
| Subjective structure | −.03 | −.27** | −.22* | −.03 | −.04 |

Note.—Since direction of relationship was predicted, one-tailed test was used for significance of correlations.

* $p < .05$.

** $p < .01$.

TABLE 3
CORRELATION (TAU) OF STRUCTURE MEASURES AND ANXIETY MEASURES WITHIN EXPERIMENTAL A AND S GROUPS

| | Anxiety Ratings | Anxiety Cards | Skin Conductance | Cue Scale I | Cue Scale II |
|---|---|---|---|---|---|
| A Group ranked on objective structure | −.32* | −.10 | .00 | −.42** | −.24 |
| S Group ranked on objective structure | −.46** | −.20 | −.12 | −.50** | −.12 |
| A Group ranked on subjective structure | −.12 | −.24 | −.28* | −.10 | .01 |
| S Group ranked on subjective structure | .13 | −.33* | −.24 | −.11 | −.02 |

Note.—One-tailed test of significance was used unless correlation was positive.
* $p < .05$.
** $p < .01$.

Structure and Anxiety Ratings and Cue Scale II, all three of which are not significantly different from zero.

Several points stand out in this table. One is that to a great extent the relationships which attained significance on the total sample (see Table 2) either attain or approach significance within the separate experimental groups. Second, the relationships do not seem equivalent in the two experimental groups examined separately. In four out of the five correlations with Objective Structure, the S group showed somewhat larger absolute negative relationships than the A group. Third, Cue Scale I showed comparatively large correlations (−.42 and −.50) with Objective Structure within both subgroups, but when the total sample was used, the correlation shrank to −.18.

The first point, concerned with correlations between structure and anxiety within groups, is consistent with the hypotheses. The within-groups analyses apparently eliminated the artificial experimental condition of instructing interviewers to behave in a structured or ambiguous way. With only the natural variation in structuring behavior the interviewers planlessly provided, the results are still of a kind suggested by the hypotheses.

As for the second point, one interpretation is that interviews in the two experimental groups necessarily had different psychological atmospheres which were related to the experimental conditions which distinguished them, and that variation in structuring behavior by the interviewer subsequently had a somewhat different effect, depending upon this more general variable. It must be remembered that there were only two real differences between the A and S groups. One was that the interviewers were instructed to be unstructured in the A group and structured in the S; the other was that they told patients early in the A interviews that they would not structure it for them, repeating this orientation several times during the interview if necessary. Since one group (A) was given to understand there would be little structuring and the other (S) had no such orientation, one may hypothesize that subsequent variability in structuring behavior by the interviewer was differently interpreted and reacted to by patients in the two groups. The finding of somewhat larger negative correlations between anxiety and structure within the S than the A group is consistent with this *post hoc* hypothesis. Seemingly, patients whose interviewers said they were going to be unstructured and instead were fairly structured showed less anxiety than did patients whose interviewers started off with a structured approach and wound up being fairly unstructured. This is made clearer by considering the third point brought out in the table, that of rather higher correlations between Cue Scale I and Objective Structure within groups than when the groups were combined.

Scatterplots of Cue Scale I and Objective Structure within A and S groups are presented in Fig. 1 side by side. The two cases of A whose Objective Structure overlapped the range of S interviews are circled. Similarly, the two S interviews whose Objective Structure was so low as to overlap the range of A interviews are circled. The lines connect mean points of five intervals in the distributions and were inserted to illustrate the regressions. One can clearly see the anxiety effects in the S interviews in which the interviewers gave comparatively little structuring, and the relatively low anxiety effect in the A interviews which were moderately structured. When all interviews were pooled for analysis, the correlation of Objective Structure and speech disruption

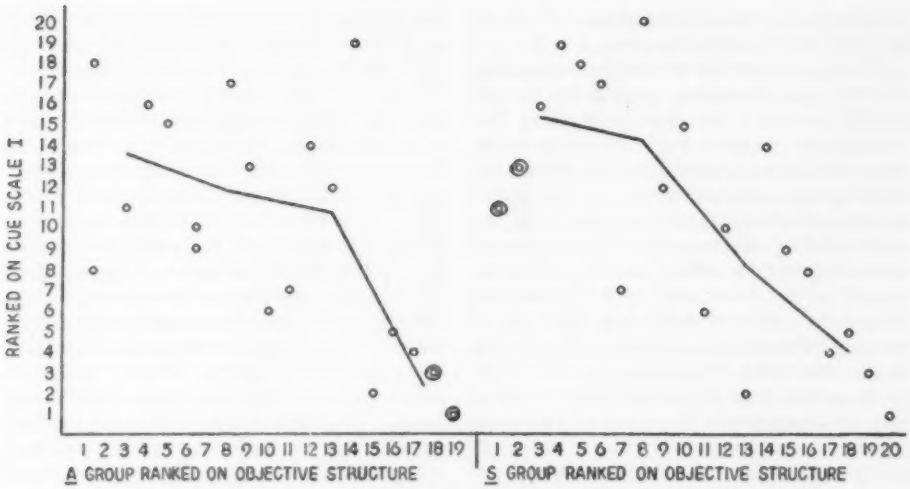


FIG. 1. SCATTERPLOT SHOWING RELATIONSHIP OF OBJECTIVE *STRUCTURE* TO CUE SCALE 1 WITHIN A AND S GROUPS

was lowered by these effects. This interpretation has a kind of common sense basis in the notion that any orientation offered to an *S* adds to the structure of the stimulus situation he faces, even the A orientation in which patients were told at the start of the interview that the interviewer would do little structuring!

The possibility of interviewer differences in arousing anxiety in the patients was explored by using the H test (16) over all interviews. None of these tests showed significance. Thus, there is no evidence that patients showed more anxiety with one interviewer than another. This factor, therefore, seems unimportant in evaluating present results.

Similarly, because disproportions in topical content during the interviews might influence findings on anxiety, an analysis was made of topics discussed during the 10-minute samples of all interviews. Interviews of A and S types were tallied on a classification of topical areas suggested by the content and subjected to chi square test. No significant differences emerged.

## DISCUSSION

A methodological factor may have had some effect on the inconsistent relationships of the anxiety measures both with each other and with the two measures of structure. Some of the measures were "global" ones, derived either from data collected during the whole interview or from impressions after the interview was completed; other measures were derived only from data obtained during the 10-minute samples, which began 20 minutes after the start of the interview. Higher relationships were found, in general, among measures of the global type or among measures of the time-sample type than when the two types were compared. In general, the global measures of anxiety (Anxiety Cards and Skin Conductance) related better to the global index of structure (Subjective Structure) than they did to the time-sample measure. Similarly, the anxiety measures which were derived from the time sample (Anxiety Ratings and Cue Scales I and II) related better to the measure of structure derived from the same sample than they did to the global measure of structure.

There are, however, some significant relationships between time-sample measures and global measures. More importantly, Objective Structure, based on consideration of the interviewer's behavior in the 10-minute samples, was found to be highly correlated with an estimate of his behavior based on the first five minutes of the interview (rho = .85), and may thus be taken as an adequate estimate of the structuredness of the whole interview. Therefore, any relationships of other measures to Objective Structure may be interpreted as

meaningful in terms of the whole interview, not just the 10-minute sample.

However, these methodological considerations do not adequately explain the lack of correlation among the anxiety measures. The two speech measures, both derived from the time sample, are uncorrelated. Similarly, Skin Conductance and Anxiety Cards, both global measures, have no significant relationship. An understanding of the lack of correlation of measures, each of which has theoretical rationale as being an index of that emotional disturbance called anxiety, sheds light on the nature of the concept of anxiety. The Anxiety Cards assess what the patient can or is willing to admit about the mental or somatic discomfort he experienced. They reflect tension of which the patient is *consciously aware*, although, of course, the degree of tension of which he is aware may be quantitatively more or less than what he indicates in his choice of Anxiety Cards. Skin Conductance, on the other hand, is one measure over which the patient has little conscious control (22). Previous studies have claimed both that *S*s are aware and unaware of changes in skin conductance (19, 32). Studies generally agree that it represents "built up" (unexpressed) tension. Perhaps there is a threshold of awareness for skin conductance changes which varies with the individual. The two uncorrelated Cue scales represent speech variables indicative of "emotionally involved" or disorganized verbal behavior. Upon analysis, some of these can be understood as direct instances of disorganization (incomplete sentence or breaking in of new thoughts) and others as defenses against or attempts to control disorganization (repetition or modification of words and phrases). Very likely, these cues are not subject to the person's awareness. The Anxiety Ratings, which are made by a clinical judge from the patient's verbal productions, can also be considered as based on a variety of cues of emotional disturbance or defense against disorganization of which the patient is essentially unaware.

The suggestion of relationship between Anxiety Cards and Skin Conductance (tau = .14) may be a reflection of the "aware" element of Skin Conductance with the "Aware-and-willing-to-admit-it" element in the Anxiety Cards measure. In any case, these low correlations force the conclusion that if all of these measures do indeed refer to anxiety, anxiety is a combination of "emotional" phenomena, each bearing little relationship to the other, some being available to the person's awareness while others, perhaps, are not although they are identifiable by an observer. Lacey and Van Lehn (18) and others (6, 21, 22) offer a theory of emotional reaction in which each individual is thought to have his own hierarchy of patterns of physiological reaction, in opposition to the view (3, 4) that certain specific emotional states evoke specific reactions in all people. Also, there is support from personality studies of psychic defense reactions that persons not only have individual hierarchies of defenses but that some use their defense system over all conflict situations, whereas others employ various reaction mechanisms, depending on the nature of the conflict situation and its relevance to the individual's unique personality dynamics (11, 12, 27, 29). This approach to the study of personality assumes that there are certain reaction mechanisms which are widely available to organisms as a group but which have varying salience, perhaps over wide temporal or situational spheres or changing with different situations, for different individuals. Applying this type of consideration to the indices of anxiety, one can imagine that each may indeed be a measure of anxiety but that individuals differ markedly in their expression of tension in speech, in autonomic changes, and in awareness of mental and somatic tension. Moreover, within the individual, there may be changes in the pattern of reaction mechanisms, depending on his threshold of tolerance for tension. Anxiety, then, may be an emotional state not subject to stable measurement by any one behavioral or physiological yardstick in any group of individuals, or in any one individual in different circumstances of stress or motivation.

Thus far, perception of structure has been considered as subject to influence by the objective nature of the situation and by the patient's anxiety. Actually, the relationship between perceived structure and patient anxiety may be a reciprocal one, the patient's report of structure being influenced to some extent by the degree to which he felt threatened by the interview. The findings in regard to Sub-

jective Structure and anxiety are more easily understood in this light. The two anxiety measures which showed correlation with Subjective Structure were Anxiety Cards and Skin Conductance. These measures are probably most representative of the patient's awareness of tension. Frenkel-Brunswik (10) found, in this connection that those who are tolerant of ambiguity (who perceive the situation as ambiguous if it objectively is) are able to tolerate the anxiety involved, while those rigid persons who perceptually engraft structure on an ambiguous situation do so to avoid anxiety. The present findings support Frenkel-Brunswik's in that perception of ambiguity was indeed related to "feeling anxious." The findings in relation to Objective Structure, however, demonstrate that people may be anxious at a level below awareness when facing an ambiguous situation. Was such the case with Frenkel-Brunswik's intolerant group?

Finally, although the goals of therapy most often involve reduction of anxiety—not elimination, but reduction to a manageable and noncrippling level (26, 30)—theorists differ as to the value of anxiety as a therapeutic agent. Analytic theorists depend on anxiety as a motivator for progress (2, 25), whereas client-centered therapists feel that anxiety reduction must be a constant goal "... because only in a non-threatening situation the individual's capacity for growth is freed" (5, p. 360). In any case, this study demonstrates that interviewers can deliberately vary the degree of structure in the situation. In view of the findings of a relationship between Objective Structure and various types of anxiety, this suggests that structuring behavior can be planfully used by a clinician as a technique for the therapeutic control of anxiety when he thinks his patient has surpassed or failed to achieve some optimal anxiety level. The research implications of this possibility are clear and important.

## SUMMARY

Considering anxiety in its meaning as a momentary state, not as a personality variable, this study tested the hypotheses that anxiety is related to the ambiguity of the presenting situation as objectively measured and as perceived by the subject.

Forty neuropsychiatric patients in the Dearborn VA Hospital were given clinical evaluative interviews by four interviewers. In half the interviews, the interviewer gave few clues to guide the patient in discussing his life and problems (Ambiguous condition). In the other interviews, the interviewer took an active part in guiding the discussion (Structured condition). All interviews were tape-recorded, and a record was obtained of patients' skin conductance throughout the interview. Patients were then questioned as to their reactions to the talks.

Two measures of structuredness were obtained: a measure of the structuring qualities of the interviewer's verbal behavior as analyzed from the typescripts, and a rating of the patient's perception of the structuredness of the interview situation as analyzed from his post-interview report. Five different measures of anxiety were used: changes in palmar skin conductance, ratings of anxiety by a clinical judge reading samples of the interview transcript, the patient's own report of tension by means of a modified check-list, and two different sets of indices of disturbed speech.

Four out of the five anxiety measures showed significant relationships with objectively measured structure, and two out of the five showed significant relationships with structure as the patient perceived it, partially confirming the hypotheses.

The general lack of correlation of the anxiety measures with each other and their patterns of relationship to the two measures of structure were discussed in terms of the units of information on which the measures were based, and in terms of the multidimensional nature of anxiety response mechanisms and the interrelationships of anxiety, perception, and personality structure. Findings of the study were discussed in relation to a theory of anxiety and to psychotherapy.

## REFERENCES

1. ADORNO, T. W. *et al. The authoritarian personality.* New York: Harper, 1950.
2. ALEXANDER, F., & FRENCH, T. M. *Psychoanalytic therapy.* New York: Ronald Press, 1946.
3. ALEXANDER, F. & FRENCH, T. M. *Studies in psychosomatic medicine.* New York: Ronald Press, 1948.
4. AX, A. F. The physiological differentiation of fear and anger in humans. *Amer. Psychologist,* 1951, **6**, 271. (Abstract)
5. BEIER, E. G. The problem of anxiety in client-

centered therapy. *J. consult. Psychol.*, 1951, **15**, 359–62.
6. CAMERON, D. E. Observations on the patterns of anxiety. *Amer. J. Psychiat.*, 1944, **101**, 36.
7. COHEN, A. R. The effects of situational structure and individual self esteem on threat-oriented reactions to power. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
8. DIBNER, A. S. Cue counting: A measure of anxiety in interviews. *J. consult. Psychol.*, 1956, **20**, 475–78.
9. DIBNER, A. S. The relationship between ambiguity and anxiety in a clinical interview. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
10. FRENKEL-BRUNSWIK, E. Personality theory and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception—an approach to personality.* New York: Ronald Press, 1951. Pp. 356–421.
11. FREUD, A. *The ego and the mechanisms of defence.* New York: International Universities Press, 1946.
12. GOLDSTEIN, K. *The organism: a holistic approach to biology.* New York: American Book, 1939.
13. HAGGARD, E. A. Experimental studies in affective processes: Some effects of cognitive structure and active participation on certain autonomic reactions during and following experimentally induced stress. *J. exp. Psychol.*, 1943, **33**, 257–84.
14. HANFMANN, EUGENIA. Psychological approaches to the study of anxiety. In P. Hoch and J. Zubin (Eds.), *Anxiety.* New York: Grune & Stratton, 1950. Pp. 51–70.
15. HILGARD, E. R. The role of learning in perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception—an approach to personality.* New York: Ronald Press, 1951. Pp. 95–121.
16. KRUSKAL, W., & WALLIS, W. Use of ranks in one-criterion variance analysis. *J. Amer. statis. Ass.*, 1952, **47**, 583–621.
17. KUBIE, L. S. A physiological approach to the concept of anxiety. *Psychosom. Med.*, 1941, **3**, 263–76.
18. LACEY, J. I., & VAN LEHN, R. Differential emphasis in somatic response to stress. *Psychosom. Med.*, 1952, **14**, 71–81.
19. LASSWELL, H. D. Verbal references and physiological changes during the psychoanalytic interview: a preliminary communication. *Psychoanal. Rev.*, 1935, **22**, 10–24.
20. LUCHINS, A. The stimulus field in social psychology. *Psychol. Rev.*, 1950, **57**, 27–30.
21. MALMO, R. B., & SHAGASS, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, **11**, 9–24.
22. MALMO, R. B., & SHAGASS, C. Physiologic study of symptom mechanisms in psychiatric patients under stress. *Psychosom. Med.*, 1949, **11**, 25–29.
23. OSBURN, H. G. An investigation of the ambiguity dimension of counselor behavior. Unpublished doctoral dissertation, Univer. of Michigan, 1951.
24. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–54.
25. REICH, W. *Character analysis.* New York: Orgone Institute Press, 1949.
26. ROSENBERG, E. Anxiety and the capacity to bear it. *Int. J. Psychoanal.*, 1949, **30**, 1–12.
27. SHIRE, A. Personality correlates of preferences among psychoanalytic mechanisms of defense. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
28. SINHA, A. Experimental induction of anxiety by conditions of uncertainty. Unpublished doctoral dissertation, Univer. of Michigan, 1950.
29. SINNETT, E. R. An experimental investigation of the defense preference inquiry for the Blacky Pictures. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
30. SUTICH, A. The growth-experience and the growth-centered attitude. *J. Psychol.*, 1949, **28**, 293–301.
31. VAN DER VALK, J. M., & GROEN, J. Electrical resistance of the skin during induced emotional stress. *Psychosom. Med.*, 1950, **12**, 303–14.
32. WOODWORTH, R. S. *Experimental psychology.* New York: Henry Holt, 1938.

Received August 29, 1956.

# THE APPLICATION OF LEARNING THEORY TO THE TREATMENT OF TICS

AUBREY J. YATES[1, 2]

*Institute of Psychiatry, Maudsley Hospital, London*

SINCE the classic work in 1907 by Meige and Feindel (19) on the treatment of tics as examples of learned habits, little progress in this field has been reported, although their results were encouraging. In 1932, however, Dunlap (5) claimed that undesirable habits (including tics) could be eliminated by making the subject repeat them voluntarily and deliberately. But Dunlap gave few details of his procedure and no experimental demonstration of its validity, merely asserting that the tics disappeared rapidly. While his technique of negative practice has been accepted (at least in so far as the treatment of tics is concerned) in at least one modern textbook of psychology (1), an adequate rationale and experimental confirmation of its utility are still lacking.

More recently, attempts have been made to derive methods for the treatment of behavioral disorders from the constructs of modern learning theory (3, 12, 13, 16, 21, 22, 25, 26).[3] This paper reports a rationale for such treatment in the case of a tiqueur and the results of experiments carried out in order to test the validity of the deductions made from the theory.

## THE TIC AS A LEARNED HABIT

The following theoretical model is proposed to show that some tics can be conceptualized as learned responses of a particular kind. The justification for proposing such a model lies in the deductions which are derived from it concerning treatment.

It is hypothesised that some[4] tics may be drive-reducing conditioned avoidance responses, originally evoked in a highly traumatic situation. In this situation, intense fear is aroused and a movement of withdrawal or aggression is made. If the movement produces or coincides with the cessation of the fear-inducing stimulus, it acquires strength through reinforcement. On subsequent occasions, through stimulus-generalization (including internal symbolization), conditioned fear ("anxiety") may be aroused, which is then reduced by the performance of the movement. In this way, the tic comes to be elicited by a large variety of stimuli and eventually achieves the status of a powerful habit.

Several points should be noticed about this model. First, although no direct experimental evidence is available to show that tics do arise in this way, two lines of investigation suggest that the formulation is a reasonable one. Wolpe (26), Solomon and Wynne (24), and others have demonstrated that animals placed in a highly traumatic situation develop conditioned avoidance responses which apparently reduce the anxiety associated with the original situation and which are highly resistant to extinction. Again, clinical investigations of the history of tiqueurs (8) have led some workers to suggest that the tic is an avoidance response arising originally in a highly traumatic situation, especially in childhood. Second, the kind of response evoked may be determined partly at least by the mode of response characteristic of the subject in any stressful situation. Malmo and his colleagues (17, 18) have presented evidence suggesting that some neurotics react to stress with cardiovascular responses, whereas others react to the same stress with muscular responses. Experiments by Lacey, Bateman, and Van Lehn (14) similarly suggest that the nature of the stressful situation is not a major variable, at least so far as autonomic functions are concerned.

The suggested model follows closely the proposed treatment, however, should be applicable to any tic which is a learned response.

[1] Now at the University of New England, Armidale, New South Wales, Australia.

[2] The author would like to thank J. Inglis and H. G. Jones for helpful discussion on the theoretical part of the paper, and Linford Rees for permission to carry out the investigation.

[3] Earlier experiments have been reviewed by Hilgard and Marquis (9), Crafts *et al.* (4), and Lehner (15).

[4] It is not suggested that tics can arise only in the way hypothesized. Some tics appear to arise from simple imitation, expecially in young children, or as an avoidance response originally arising from some irritating but nontraumatic stimulus such as a tight collar. The

two-factor theory of learning expounded by Mowrer (21). In terms of Hullian (10) learning theory, the reaction potential of the tic at a given moment may be conceived as a multiplicative function of the habit strength ($_sH_R$) of the tic (determined mainly by the number of times it has previously been evoked) and the momentary drive strength of anxiety (D), which fluctuates from time to time. Since habit strength increases as a simple negatively accelerated positive growth function and eventually reaches an asymptote, further performance of the tic cannot increase its habit strength beyond a given point.

### DERIVATION OF METHOD OF TREATMENT

According to the model, then, the tic may be treated as a simple learned habit which has attained its maximum habit strength. In terms of the theory, it should be possible, therefore, to extinguish the habit by building up a negative or incompatible habit of "not performing the tic." If the subject is given massed practice in the tic, then reactive inhibition ($I_R$) should build up rapidly.[5] When $I_R$ reaches a certain critical point, the patient will be forced to "rest" or not perform the tic. This habit ($_sI_R$) of not performing the tic will be associated with drive-reduction due to the dissipation of $I_R$ and hence will be reinforced. With repeated massed practice, therefore, a negative habit ("not doing the tic") will be built up, incompatible with the positive habit of doing the tic. Furthermore, the repeated voluntary evocation of the tic should *not* serve to increase the habit strength of the tic, since it is already asymptotic and consequently not subject to strengthening by massed practice.

At any given moment, therefore, the effective reaction potential of the tic (i.e., the *ability* of the patient to perform the tic) will be a resultant of the forces in the equation

$$_s\overline{E}_R = (_sH_RXD) - (I_R + {}_sI_R)$$

As $_sI_R$ increases, $_s\overline{E}_R$ should diminish to a point at which the excitatory and inhibitory parts of the equation are equal. The response tendency of the tic will then be below threshold level, and behaviorally, extinction should be complete. During this process, of course, $_s\overline{E}_R$ would be expected to show fluctuations, since it is affected by a number of factors (notably drive level) which are uncontrolled ($_sO_R$). Any increase in drive may, of course, temporarily mask the effects of an increase in $_sI_R$. It is interesting to note that Calvin *et al.* (2) recently produced evidence to show that the continued evocation of a habit under conditions of massed practice in rats eventually leads to its extinction, although the process may be a slow one.

The optimum conditions of massed practice under which conditioned inhibition ($_sI_R$) may be expected to grow most rapidly and effectively are unknown with respect to the extinction of tics. The present experiments represent, therefore, an attempt to discover these conditions.

### THE PATIENT

The *S* was a female psychiatric patient, 25 years old, who was highly cooperative and apparently had a strong desire to get well. She was of high average intelligence, and was markedly neurotic and slightly more extroverted than the average normal as measured by the *Guilford-Martin Personality Inventory of Factors S, T, D, C, R*.[6] Four clear-cut tics were chosen for the experiment: a complex stomach-contraction breathing tic, a nasal "explosion" (expiration), a coughing tic, and an eye-blink tic. These tics appeared to have started originally following two very traumatic experiences about ten years previously when *S* felt that she was being suffocated while undergoing anaesthesia; she said she was terrified that she was going to die and struggled madly. She could not bear the thought of an anaesthetic mask and could not tolerate any object's being placed over her face. Superficially, at least, these tics seemed to be cond..ioned avoidance responses originally established in a traumatic situation.[7] She also said she felt a need to do the tics and experienced relief when they occurred (i.e., they were drive-reducing). They varied mark-

[5] With massed practice the tic will also, for the most part, cease to be reinforced, because, since anxiety is only intermittent in *S*, it will be unaccompanied by drive reduction.

[6] Eysenck (7) has shown that these scales are reasonably good measures of neuroticism (C and D scales) and introversion-extraversion (R scale).

[7] Presumably, escape in such a situation from the noxious stimulation was effected by the patient's becoming unconscious.

TABLE 1
CHANGES IN MEAN FREQUENCY OF FOUR TICS UNDER CONDITIONS OF VOLUNTARY EVOCATION

| Stomach | | | Eyeblink | | | Nose | | | Throat | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sessions | M | $\sigma$ | Sessions | M | $\sigma$ | Sessions | M | $\sigma$ | Sessions | M | $\sigma$ |
| 1–50 | 52.83 | 4.06 | 1–50 | 36.96 | 4.74 | 1–50 | 45.56 | 5.41 | 1–50 | 42.00 | 4.17 |
| 51–100 | 40.54 | 3.90 | 51–100 | 31.96 | 2.48 | 51–100 | 35.11 | 3.02 | 51–100 | 32.18 | 3.55 |
| 112–150 | 41.69 | 2.58 | 101–150 | 31.30 | 1.63 | 112–150 | 32.06 | 3.13 | 101–150 | 31.52 | 3.53 |
| 151–175 | 19.40 | 2.15 | 151–200 | 26.87 | 2.00 | 151–180 | 23.71 | 2.73 | 151–200 | 23.91 | 2.25 |
| 176–200 | 15.26 | 1.57 | | | | | | | | | |
| 201–220 | 11.95 | 1.09 | 201–220 | 26.04 | 1.60 | 201–220 | 18.97 | 4.25 | 201–220 | 20.27 | 3.33 |
| 221–285 | 9.67 | 1.04 | 221–285 | 20.34 | 1.74 | 221–248 | 13.23 | 2.12 | 221–285 | 15.51 | 2.10 |
| 286–315 | 8.11 | 0.75 | 286–315 | 21.14 | 1.48 | 286–315 | 8.05 | 1.90 | 286–315 | 12.88 | 1.73 |

edly from time to time in frequency of occurrence. It was decided to treat the four tics concurrently but independently.

## EXPERIMENT I

### *Sessions 1–100*

The purpose of this experiment was to test the general hypothesis that massed practice leads to a significant decrement in the ability of *S* to respond voluntarily.

Two sessions per day were carried out, each consisting of five one-minute trials for each tic under conditions of massed practice, with one minute's rest between each set of five trials.[6] *S* was instructed to reproduce each tic as accurately as possible and to repeat it without pause during the practice periods. No stress was laid on speed. A complete session lasted exactly 45 minutes. One such session was carried out at the hospital each day under supervision; *S* carried out the other on her own. The order of practising the tics was varied at random. The score recorded was the number of tics per minute as counted by *E* in the case of the nasal and throat tics and by *S* in the case of the stomach and eyeblink tics. This procedure was followed for 100 sessions (50 days).

The frequency of occurrence of each tic per minute (under test conditions) almost invariably showed a steady decline from trials one to five within a single session. It was considered that the average of the five trials per session would give a reliable estimate of the strength of the tic under conditions of voluntary evocation in any particular session. The mean score for the first 50 sessions was compared with the mean score for the second 50 sessions for each tic separately.

The results (Table 1) are remarkably consistent for the four tics. The mean score for the second set of 50 sessions is significantly lower than the mean for the first set of 50 for each tic (values of $t$ were as follows: stomach, 15.42; eyeblink, 6.60; nose, 11.93; throat, 12.68; $p$ was $< .001$ in each case). The results, therefore, are in accordance with prediction and offer support for the general theory.

## EXPERIMENT II

### *Sessions 101–150*

Three hypotheses were tested over Sessions 101–150.

*Hypothesis I.* It seemed reasonable to contend that further massed practice under standard conditions would produce a further decline in the tic frequency. To test this hypothesis, the eyeblink and throat tics were practised under standard conditions.

*Hypothesis II.* The difference between means of the first and second sets of 50 sessions in Experiment I might be due to $I_R$ not completely dissipating between sessions. To test this hypothesis, practice on the stomach tic was discontinued for 11 sessions; during the ensuing Sessions 112–150, standard conditions were resumed. If the decline in Experiment I were due to $I_R$ alone, the mean for Sessions 112–150 should not be lower than that for Sessions 1–50, i.e., recovery of function should take place.

*Hypothesis III.* To test the notion that the growth of $_sI_R$ would be facilitated by increasing the duration of massed practice, the nasal tic was practised for 15 min. daily without a break under instructions to keep the amplitude as strong and as constant as possible. This procedure apparently had the desired effect of increasing the number of involuntary rest pauses, a crucial factor in the growth of $_sI_R$, since in the last few minutes of the period, the rate of responding typically dropped to 3–4 per minute. Five sessions (one per day) were carried out in this way, the prolonged period always following practice on the other tics to avoid interference with them. No standard practice on the nasal tic was carried out at this time. During Sessions 112–150, standard conditions were imposed. It was predicted that the mean for Sessions 112–150 would be significantly below that for Sessions 51–100.

The results of this experiment are reported in Table 1. Hypothesis I was not confirmed. For both eyeblink and throat tics, the mean frequency per session was not significantly below that for Sessions 51–100; (for the eyeblink, $t = 1.570$, $p = > .05$; for the throat, $t = 0.927$, $p = > .05$). However, no recovery of function took place, since in both cases the mean for Sessions 101–150 was significantly lower than that for Sessions 1–50; (for the eyeblink, $t = 7.982$, $p = < .001$; for the throat, $t = 13.563$, $p = < .001$).

Hypothesis II was not confirmed. The mean for Sessions 112–150 was significantly lower than the mean for Sessions 1–50 ($t = 14.928$, $p = < .001$). However, the mean for Sessions 112–150 was not significantly different from the mean for Sessions 51–100, ($t = 1.589$, $p = > .05$).

Hypothesis III was confirmed. The mean for Sessions 112–150 was significantly lower than the mean

[6] Designated henceforward the standard procedure.

for Sessions 51–100, ($t = 4.655$, $p = < .001$). Although the decline was not very large, it should be remembered that only five sessions of increased massed practice were given.

During this experiment, therefore, *S* seemed to be on a plateau with respect to three of the tics, but recovery of function did not take place, even when practice was entirely discontinued for a week. Further, it was apparent that the growth of ${}_{s}I_{R}$ could be facilitated in some degree by increasing the amount of massed practice.

## EXPERIMENT III

### *Sessions 151–200*

Three hypotheses were again tested over Sessions 151–200.

*Hypothesis I.* Practice of the eyeblink and throat tics was continued under standard conditions on the assumption that during Sessions 50–150 *S* had been on a plateau with respect to these tics; it was predicted that further practice would eventually induce a decline.

*Hypothesis II.* It seemed probable that within the standard practice conditions, the intensification of the response would increase the rate of decline. Dunlap (5) stressed the importance of instructing *S* to pay careful attention to the tic. *S* was, therefore, asked to reproduce the stomach tic as precisely and as intensively as possible. The conditions of practice were otherwise standard.

*Hypothesis III.* The prolongation of the period of practice for the nasal tic was continued. However, the method was varied. Each day, at the ***end*** of the hospital session only, *S* did 15 min. massed practice of the nasal tic as in Experiment II. In the second session of that day, the nasal tic was not practised (since the $I_R$ associated with the 15-min. practice period might not have wholly dissipated and could affect performance). However, at the ***beginning*** of the hospital session each day, the nasal tic was practised under standard conditions. This served as a measure of the influence of the previous day's intensive practice. The standard nasal practice was followed by the throat and eyeblink tics, which always preceded the stomach tic to avoid the effects of fatigue induced by that tic under intensive conditions. In this way, at least 12 hours elapsed between the intensive and the standard nasal practises. During this experiment, there were 20 intensive and 30 standard practice periods for the nasal tic.[9]

The results, as shown in Table 1, supported Hypothesis I. The mean for Sessions 151–200 was significantly lower than the mean for Sessions 101–150 for both eyeblink and throat tics, (for the eyeblink, $t = 12.137$, $p = < .001$; for the throat, $t = 9.931$, $p = < .001$).

With regard to Hypothesis II, the results obtained with the stomach tic were not directly comparable with those obtained in Sessions 1–150 because of the altered instructions within the standard conditions. When, however, Sessions 151–175 were compared with Sessions 176–200, a significant decline was apparent in the latter set, ($t = 7.793$, $p = < .001$). The hypothesis may, therefore, be regarded as supported.

Hypothesis III was also supported. Extensive massed practice led to a striking decline in frequency, the mean for Sessions 151–180 being significantly below that for Sessions 112–150 ($t = 11.608$, $p = < .001$).

The results of Experiment III suggested, therefore, that the standard procedure led to a continuing decline in frequency, that extensive massed practice had a very striking effect on frequency of response, and that increasing the intensity of response within the standard situation also led to a decline in frequency.

## EXPERIMENT IV

### *Sessions 201–220*

The purpose of Experiment IV was to test more rigorously the hypothesis that the decline in frequency of the four tics was a function of the growth of ${}_{s}I_{R}$ rather than of the accumulation of $I_R$. When practice on the stomach tic was discontinued in Experiment II, *S* later reported that she had been unable to inhibit it during the 15-min. practice of the nasal tic. Hence, the earlier experiment seemed inconclusive. *S* was, therefore, given a two-week holiday during which no practice of any kind was carried out. It was predicted that no recovery of function would take place. Following this interlude, the four tics were practised under standard conditions only, for 20 sessions.

The results, as shown in Table 1, indicated that the eyeblink tic showed no change following complete rest, ($t = 1.00$, $p = > .05$). The other three tics all showed a significant further decline during Sessions 201–220 compared with the level reached in Experiment III (for the stomach, $t = 8.37$, $p = < .001$; for the nose, $t = 3.98$, $p = < .001$; for the throat, $t = 4.49$, $p = < .001$). The decline was particularly striking in the case of the nasal tic. The hypothesis that no recovery of function would occur seemed, therefore, to be supported, but the significant decline following rest in the case of three tics remained to be explained.

## EXPERIMENT V

### *Sessions 221–275*

Three hypotheses were tested in this experiment.

*Hypothesis I.* Massed practice under standard conditions was continued for the eyeblink and throat tics, the prediction being that a further significant decline would result.

*Hypothesis II.* Massed practice under standard conditions but with intensification of the response was continued for the stomach tic, the prediction being that a further significant decline would result.

*Hypothesis III.* It was decided to test whether the rate of growth of ${}_{s}I_{R}$ could be significantly increased by an increase in the ***rate*** and ***intensity*** of repetition of the nasal tic. To test this hypothesis, the 15-min. period of massed practice was used. This period was divided into periods of 5 seconds of massed practice followed by 10 seconds of rest. During the 5 seconds, *S* performed the nasal tic as rapidly and as intensively as possible. It was expected that $I_R$ would accumulate more rapidly and would dissipate less between indi-

[9] The numbers are unequal because *S* was later allowed to do the 15 min. practice on her own on weekends.

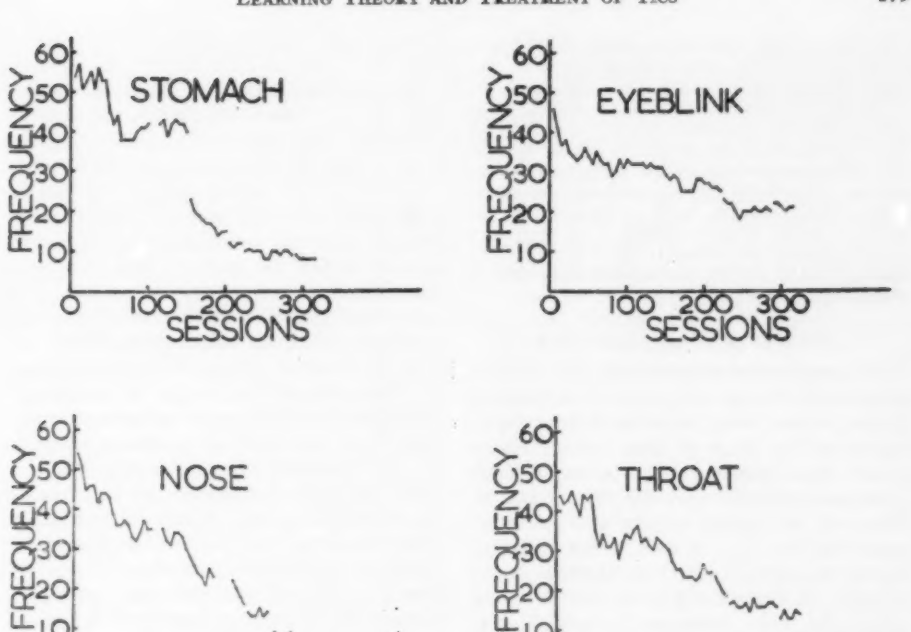


FIG. 1. CHANGES IN FREQUENCY OF FOUR TICS VOLUNTARILY EVOKED UNDER VARIOUS CONDITIONS OF MASSED PRACTICE

vidual tics. Because 10 seconds of rest would thus be expected to have a high reward value, ${}_{S}I_{R}$ should be more strongly reinforced. Apart from the change in procedure for the 15 minutes of practice, the method adopted was identical with that described in Experiment III.

The results, as shown in Table 1, indicated that all three hypotheses were confirmed[10] (for the eyeblink, $t = 13.67$; for the stomach, $t = 8.26$; for the nose, $t = 5.21$; for the throat, $t = 6.16$; $p$ was $< .001$ in each case).

## EXPERIMENT VI

### *Sessions 276–315*

The purpose of this experiment was to test the hypothesis that a very prolonged period of massed practice followed by a prolonged rest would lead to a significant decrement in response. It had been found that in Sessions 201–220 a highly significant decline took place on three of the four tics following two weeks' rest, the decline being most striking in the case of the nasal tic. The possibility arose, therefore, that the crucial set of conditions for increasing ${}_{S}I_{R}$ rapidly lay in a combination of very intensive practice followed by a long rest. In general, it was hypothesized that optimum increments of ${}_{S}I_{R}$ would be obtained if sufficient massed practice were given to allow the accumulation of a maximum amount of $I_R$, followed by a rest period sufficient to allow the complete dissipation of $I_R$. To test this hypothesis, four one-hour sessions were given, during which uninterrupted massed practice of the nasal tic was carried out. The other tics were not practised during these sessions, each of which was separated from the next by a three-day interval. There were no standard practices of the nasal tic during this period. The other tics were practised under the usual conditions during the second daily session only (i.e., once per day, constituting Trials 276–285). At the end of this period, the patient was given a three-week holiday during which no practice of any kind was carried out. Following this rest, 30 standard sessions (286–315) were given for all four tics.

The results, as shown in Table 1, indicate that the nasal tic showed a very significant decline in the mean for Sessions 286–315 compared with Sessions 221–248 immediately preceding the four one-hour practices ($t = 7.99$, $p = < .001$). Throat and stomach tics showed a significant but much smaller decline (for the throat, $t = 6.43$, $p = < .001$; for the stomach,

[10] Table 1 shows the means for Sessions 221–285 for eyeblink, throat, and stomach tics. Sessions 276–285 were, however, actually performed for these tics during Experiment VI while the nasal tic was being given four one-hour practices.

$t = 8.32$, $p = < .001$). The eyeblink tic showed a small but significant recovery ($t = 2.32$, $p = .02$). Thus, the hypothesis was generally supported.

At the conclusion of this experiment, the mean score for the last 30 sessions (286–315) was compared with the mean score for the first 30 sessions (1–30).[11] The percentage decrement in the frequency of voluntary responding for each tic was as follows: nasal, 84.14; throat, 68.83; stomach, 63.42; eyeblink, 51.47. The course of decline is shown in Fig. 1, in which each point represents the average of five sessions (i.e., 25 one-min. trials). Gaps in the curves indicate rests or changes in treatment.

### CHANGES IN INVOLUNTARY TICS

It is notoriously difficult to obtain a reliable estimate of the rate of involuntary responding in tics, which fluctuates markedly from day to day, even from hour to hour. While a *change* in the rate of response would be very difficult to assess reliably, a complete cessation of involuntary responding outside the test situation would be apparent to *S*, to her acquaintances, and to *E*. *S* had not reached this clinical criterion at the end of Experiment VI. At the end of the series, however, she was asked to write a frank account of her experience over the period of treatment, both in relation to the tics and to her general adjustment.

Her report indicated that the nasal tic "has almost vanished. It returns occasionally . . . every now and then . . . but most definitely it does not plague me every few minutes as before. It is absent for days on end." The eyeblink and throat tics were reported as sharply reduced in frequency. The stomach tic, however, had changed very little. She also reported that she had "improved in all departments generally," was "able to relax in a bus or train, able to read or execute a crossword puzzle." There was continual reference to the fact that not doing the movement did not leave her frustrated. She concluded that she was, "not tired of the treatment, if anything now that I have a taste of success the desire is greater." She also pointed out that "one of the striking features is that my closest friends notice the change." The specific and general changes were confirmed by her physician.[12]

[11] For the stomach tic, the comparison was between Sessions 151–180 and 286–315.

[12] During these experiments, *S* received no other treatment but saw her physician for a few minutes about once a month.

### DISCUSSION

The results obtained in these experiments support the theory proposed to explain the genesis of certain tics and the deductions made from it regarding a rational method of treatment. Within the test situation there was objective evidence of a striking decline in the rate of voluntary responding for all tics. The rate of decline was most striking in that tic (nasal) which received the largest amount of massed practice. There was no evidence of any recovery of function following prolonged rest. The data further suggest that the optimum conditions for the production of conditioned inhibition may be a period of very prolonged massed practice, leading to the accumulation of large amounts of reactive inhibition, followed by a very prolonged rest to allow for the total dissipation of reactive inhibition. There was subjective evidence suggesting that conditioned inhibition generalized beyond the test situation itself and that most generalization occurred with respect to the tic which had received most massed practice. It is not suggested, however, that the results outlined above can be unequivocally interpreted. Under clinical conditions of testing and with limited time available, perfect experimental control and balance of design proved impossible to obtain. Thus, it is possible that the decline in Sessions 286–315 in the nasal tic was a function entirely of the amount of massed practice given and not of the combination of practice and rest. Although no standard practices were carried out on the nasal tic between the four one-hour periods, *S* reported that she became much worse outside the test situation during this time. However, during the three weeks' rest period which followed, she reported that she rapidly lost the tendency to perform the tic involuntarily, which suggests that the rest was an essential part of the treatment. These results imply that the cumulative reinforcement associated with the dissipation of reactive inhibition may continue for longer periods of time than is commonly supposed. This interpretation is consistent with the findings of Jahnke and Duncan (11).

A number of considerations which arose during the experiments merit brief mention. It would have been useful to have left one tic untreated and to have taken control read-

ings under standard conditions at regular intervals. Under these conditions, little or no conditioned inhibition would be expected and the prediction would be that the untreated tic would not decline during the course of the experiment. Second, an objective measure of the frequency of each tic outside the test situation at the beginning and end of the experiments is highly desirable. It should not be too difficult to measure the frequency of tics reliably in a natural situation. Third, attention should be paid to the amplitude as well as the frequency of the tics; a measure of both variables would be useful. Fourth, *S* should be instructed from the start to reproduce the tic as accurately as possible in order to facilitate the growth of $I_R$.

Nevertheless, certain features of the obtained results are highly suggestive. The shape of the curves suggests, in some respects at least, the obverse of a learning curve. There is no evidence of an initial rise in frequency, indicating that the habit strength of the tic was probably at its asymptote before massed practice was begun. Again, the decline in frequency appears to proceed according to the usual learning process. There are periods of stability (plateaus) followed by sudden falls, suggesting that a negative habit was actually being learned. This finding agrees with the results of Edmonson and Amsel (6), who found that in the extinction of conditioned avoidance responses, long periods during which no change took place, alternated with sudden sharp declines. The curves also show the sessional fluctuations common in learning curves, where all the relevant factors (notably, in this instance, drive level) cannot be controlled.

An important problem concerns the relationship between the anxiety and the tic. *S* reported at the end of treatment that she was much less anxious than formerly (reflected in many particular examples of behavior she gave). This reduction in anxiety follows directly from the theory. If the ability to respond is significantly reduced, then the anxiety leading to the tics will persist. However, since no traumatic event happens in fact as a result of the anxiety, the latter, being itself a conditioned response, gradually extinguishes. This formulation suggests an alternative method of treatment. If the tics are prevented from occurring, then reduction of the anxiety should follow. When this happens, the tics will also decline, since they are a response to anxiety. Meige and Feindel (19) reported cases in which paralysis of facial tics for several days led to their disappearance for a period of several months. This hypothesis also receives support from the findings of Solomon and his colleagues (23) that conditioned avoidance responses can be extinguished in dogs by presenting the CS a large number of times but preventing the avoidance response by restraint. The usual extinction procedures, which formerly had been unsuccessful, could then be applied. These arguments are the principal justification for symptomatic treatment. It is interesting to note that they follow directly from Mowrer's two-factor theory. Mowrer himself has strongly opposed symptomatic treatment (21, p. 620) although making use of it in the treatment of enuresis. In terms of the theory, it would also be predicted that this method of treatment would be most successful with children because ${}_sH_R$ would presumably be at less than asymptotic strength. Thus, as soon as negative practice began, most of the repetitions of the tic would be contributing towards the growth of ${}_sI_R$ rather than ${}_sH_R$, and the growth of the negative habit (not-doing-the-tic) would proceed more rapidly than the growth of the positive habit.

## SUMMARY

A theoretical model was proposed to show that some tics may be conceptualized as drive-reducing conditioned avoidance responses, originally evoked in a traumatic situation.

From this model, a method of treatment was derived. It was predicted that if the tics were evoked voluntarily under conditions of massed practice, a negative habit of "not doing the tic" should be built up, resulting ultimately in the extinction of the tics and that this extinction should generalize beyond the test situation.

The results of a number of experiments support the validity of the theory. There was a significant decline in the ability to respond voluntarily under various conditions of massed practice and rest.

The optimum condition for the growth of

the negative habit appeared to be the combination of very prolonged massed practice followed by prolonged rest.

Subjective reports of the patient indicated considerable clinical improvement outside the immediate test situation.

## REFERENCES


1. BORING, E. G., LANGFELD, H. S., & WELD, H. P. *Foundations of psychology*. New York: John Wiley, 1948.
2. CALVIN, A. D., CLIFFORD, L. T., CLIFFORD, B., BOLDEN, L., & HARVEY, J. Experimental validation of conditioned inhibition. *Psychol. Rep.*, 1956, **2**, 51–56.
3. CHERRY, C., & SAYERS, B. MCA. Experiments upon the total inhibition of stammering by external control and some clinical results. *J. psychosom. Res.*, 1956, **1**, 233–246.
4. CRAFTS, L. W., SCHNEIRLA, T. C., ROBINSON, ELSA E., & GILBERT, R. W. (Eds.). *Recent experiments in psychology*. New York: McGraw Hill, 1950.
5. DUNLAP, K. *Habits: Their making and unmaking*. New York: Liveright, 1932.
6. EDMONSON, B. W., & AMSEL, A. The effects of massing and distribution of extinction trials on the persistence of a fear-motivated instrumental response. *J. comp. physiol. Psychol.*, 1954, **47**, 117–123.
7. EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. di Psicol.*, 1957, **50**, 113–140.
8. GERARD, MARGARET. The psychogenic tic in ego development. In Anna Freud, *et al.* (Eds). *Psychoanalytic study of the child*. Vol. 2. New York: Inter. Univer. Press, 1946, Pp. 133–162.
9. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning*. New York: D. Appleton-Century, 1940.
10. HULL, C. L. *Principles of behavior*. New York: Appleton Century, 1943.
11. JAHNKE, J. C., & DUNCAN, C. P. Reminiscence and forgetting in motor learning after extended rest intervals. *J. exp. Psychol.*, 1956, **52**, 273–282.
12. JONES, E. L. Exploration of experimental extinction and spontaneous recovery in stuttering. In W. Johnson (Ed.), *Stuttering in children and adults*. Minneapolis: Univer. of Minnesota Press, 1955. Pp. 226–231.
13. JONES, H. G. The application of conditioning and learning techniques to the treatment of a psychiatric patient. *J. abnorm. soc. Psychol.*, 1956, **52**, 414–419.
14. LACEY, J. I., BATEMAN, D. E., & VAN LEHN, R. Autonomic response specificity. *Psychosom. Med.*, 1953, **15**, 8–21.
15. LEHNER, G. F. J. Negative practice as a psychotherapeutic technique. *J. gen. Psychol.*, 1954, **51**, 69–82.
16. LIVERSEDGE, L. A., & SYLVESTER, J. D. Conditioning techniques in the treatment of writer's cramp. *Lancet*, 1955, **2**, 1147–1149.
17. MALMO, R. B., & SHAGASS, C. Physiologic study of symptom mechanisms in psychiatric patients under stress. *Psychosom. Med.*, 1949, **11**, 25–29.
18. MALMO, R. B., SHAGASS, C., & DAVIES, F. H. Symptom specificity and bodily reactions during psychiatric interview. *Psychosom. Med.*, 1950, **12**, 362–367.
19. MEIGE, H., & FEINDEL, E. *Tics and their treatment*. London: S. Appleton, 1907.
20. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*, New York: John Wiley, 1951. Pp. 435–472.
21. MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald Press, 1950.
22. RAYMOND, M. J. Case of fetishism treated by aversion therapy. *Brit. med. J.*, 1956, **2**, 854–857.
23. SOLOMON, R. L., SIDD, J. J., WATSON, P. D., & BLACK, A. H. The use of d-tubocurarine in the extinction of fear in dogs. *Amer. Psychologist*, 1955, **10**, 395. (Abstract)
24. SOLOMON, R. L., & WYNNE, L. C. Traumatic avoidance learning: the principles of anxiety conservation and partial irreversibility. *Psychol. Rev.*, 1954, **61**, 353–385.
25. WISCHNER, G. J. Stuttering behavior and learning: A preliminary theoretical formulation. *J. Speech Dis.*, 1950, **15**, 324–335.
26. WOLPE, J. Experimental neuroses as learned behavior. *Brit. J. Psychol.*, 1952, **43**, 243–268.


Received October 10, 1956.

# GSR CONDITIONING IN ANXIETY STATES, NORMALS, AND CHRONIC FUNCTIONAL SCHIZOPHRENIC SUBJECTS[1]

EDMUND S. HOWE[2]

*University College, London*

In recent years, increasing attention has been given to experimental studies of anxiety and conditioning and to interpathological explorations of conditioning. There is now general agreement that anxiety, assessed either psychiatrically or psychometrically, plays a part in human eyelid (22, 23, 24, 25) and GSR conditioning (2, 26, 27). Several studies have now confirmed that the neurotic *S* conditions more readily in a threat situation than does the normal (23, 25), a difference which may be explained by an increase in motivation derived from anxiety. The data on the conditionability of psychotics are, on the other hand, rather scanty and unsystematic (1, 15, 19, 21). Two recent studies (23, 25) have reported data and deductions for psychotics with the conclusion that these *S*s condition more readily than both neurotic and normal *S*s. This conclusion is not consistent with the findings of Paintal (18), for example, who examined GSR to electric shock and to threat of shock in 450 normal and 450 psychotic *S*s. Although there was little difference between the *S*s' response to the shock, there was a significantly smaller response in psychotic *S*s to the *threat* of shock. The present study examined conditioned GSR behavior in hospitalized anxiety states, normals, and hospitalized chronic functional schizophrenics, on the hypothesis that these three groups involve, respectively, a descending order of total drive strength (Hull's *D*) with respect to such a nociceptive stimulus as electric shock.[3]

The question of differential drive strengths in the three subject groups may, of course, be approached empirically by a comparison of conditioning measures, but this procedure is not necessarily the optimal one. From the standpoint of Hullian theory (11), provision is made for a multiplicative combination of other drives (e.g., anxiety) with the relevant drive energizing the habit involved. In the present case, *D* may therefore be considered a direct function of anxiety level. In the case of the anxiety states, this contention has repeatedly proved fruitful (22, 23, 24, 25); i.e., there is a sound basis for predicting that the anxiety cases will show more rapid, stronger conditioning than normal controls. There is, however, a difficulty of relating Hull's *D* to the anxiety level of schizophrenics because, while such patients may well show an absence of *manifest* anxiety symptoms (leading to a prediction of slower conditioning), it may be reasonably contended (e.g., 23, 25) that their withdrawal is a defense against high (hypothetical) *internal* anxiety (leading to a prediction of more rapid conditioning). Perhaps this

[1] This paper contains a portion of the data from a thesis submitted to The Department of Psychology, University College, in candidacy for the degree of doctor of philosophy in the University of London. The writer is indebted especially to John W. Whitfield, who throughout supervised the research, for his encouragement and advice; and to Roger W. Russell and Hans J. Eysenck, each of whom made valuable comments at different stages of the work. Many thanks are also due to the members of the psychiatric, psychological, and nursing staffs at St. Ebbas Hospital, Epsom, and The Belmont Hospital, Sutton, Surrey, for their willing and continuous cooperation in making it possible for the writer to execute this study.

[2] Now with the Department of Psychiatry, University of Maryland School of Medicine.

[3] A second but major purpose of the study was to examine the effect of drive level *D* upon the steepness of the gradient of stimulus generalization, a relationship that has been explored by more recent writers (4, 20). This is the essential reason for extinguishing one-third of each group to a different tone. When gradients of generalization over three test points were examined at each extinction trial for each group, however, only two gradients showed a significant departure from zero slope, which could, of course, occur by chance. Because of this result, data and discussion on generalization have been entirely omitted from this report. It was reasoned *ex post factum* that the outcome concerning the gradients was a function mostly of *S*s having been given minimal familiarity with stimuli to be employed during the pretesting and conditioning phases (which would *decrease* responsiveness to stimulus differences and hence spuriously *increase* generalization). It was thus not possible to offer conclusive evidence about the hypothesis set up: namely, that amount of generalization is a function of drive strength, rather than simply of stimulus similarity (see, for example, 11, 16, 17, 20).

problem has been turned into an either/or question, neglecting the possibility that the *degree* of chronicity and deterioration in schizophrenia at least partly determines anxiety level and hence reactivity in a threat situation, the relationship probably being an inverse one.

It can thus be argued, as Malmo *et al.* concluded, that, "the much lower reactivity of the chronic schizophrenic is evidence for a trend toward lowered reaction to pain as schizophrenia progresses toward chronicity" (14, p. 63). Malmo and his collaborators suggested, however, that this interpretation is probably applicable to rather specific experiences like pain, and may not be true at all for other situations involving overt adjustments. The characteristics of chronic schizophrenics, as presented by such writers as Hoagland (6, 7, 8), Hoskins (9), Hill *et al.* (5), and others, show a lowered reactivity to equilibrium functions, central disturbances rather than peripheral ones, and inadequate responses to stress as measured by lymphocyte counts and 17-ketosteroid output. It has also been suggested by Hoskins (9) that an inadequate adrenal cortical response, rather than a higher stress threshold, is responsible for the lowered reactivity of the chronic schizophrenic. The general picture emerging from a survey of this kind of work leads to the following conceptualization of the chronic deteriorated schizophrenic: stress-produced responses which seem to be justified biologically frequently do not occur at all, or occur late, or occur to a diminished degree. This general statement is, of course, quite the converse from and at the same time complementary to what is expected in the anxiety neurotic, despite the fact that different pain thresholds have not been demonstrated for the different *S*s. With such considerations in mind, it was considered most reasonable to hypothesize low drive strength with respect to electric shock in the schizophrenic population used in this study, and to predict weaker, slower conditioning.

## METHOD

### *Subjects*

A total of 60 normal control *S*s was used, half drawn from University of London students and half from the nursing staffs of the two cooperating hospitals. Student *S*s ranged from 19 to 24 years; half of them were male, sixty per cent taking honors in psychology and the remainder coming from other departments. The sample of nurses contained an equal number of males and females, with an age range from about 18 to 25. All control *S*s were volunteers.

The anxiety *S*s were in-patients drawn from Belmont Hospital, Surrey, a National Rehabilitation Center. A total of 60 *S*s, psychiatrically diagnosed as acute or chronic cases of severe anxiety, were distributed as follows: anxiety reactions, 50; phobics, 5; compulsives, 3; hysterics, 2. There were 29 males and 31 females. All *S*s reported upon cooperated voluntarily. The ages of these patients were not recorded, but the mean was about 30 with a range of approximately ten years.

The schizophrenic group was drawn from voluntary in-patients at St. Ebbas Hospital, Surrey, and was selected on the basis of psychiatric diagnosis. A criterion of chronicity was set at two years or more of illness; but in fact the mean duration of illness was over five years. After elimination of unusable data (due, for example, to some *S*s' quitting midway through the conditioning procedure), 60 *S*s psychiatrically diagnosed as chronic, functional schizophrenic were distributed as follows: simple schizophrenia, 3; paranoid schizophrenia, 4; catatonia, 10; hebephrenia, 1; and "mixed schizophrenic symptoms," 42. There were 26 males and 34 females, with a mean age of 32 years. Patients were assigned for conditioning by the psychiatric staff. Any *S* who attempted to leave the experimental room was permitted to do so.

A further group of 36 schizophrenic *S*s proved to be unconditionable according to one of three criteria: no GSR whatsoever to twice the standard intensity of shock usually delivered to other *S*s during pretesting; complete sensory adaptation to shock (i.e., no GSR) either before or at the last conditioning trial; or zero GSR to the first extinction trial. A comparison of these *S*s with those who were conditioned is presented in the results.

### *Apparatus and Method of Recording GSRs*

The Féré method, involving impressed current, was used in the circuit described by Lacey and Siegel (13). GSRs were read from a Pye "Scalamp" galvanometer, the period of which is two seconds, and which can be read quite accurately to 0.1 mm. In order to determine *S*'s resistance, it was simply necessary to adjust $R_2$ until approximately 60 microamperes were passing through him. Since the potential drop across *S* was always constant, basal resistance was obtainable from a specially calibrated dial, superimposed upon $R_2$. The ohmic magnitude of any single GSR was thus directly proportional to the magnitude of the spot deflection. The GSR electrodes consisted of two highly polished brass discs, one inch in diameter, each firmly embedded in the base of a vacuum tube ground down to a depth of about one-half inch. Felt plugs, soaked in two per cent saline before use with each *S*, were wedged in the cups on top of the electrodes. The electrodes were firmly attached with elastic bandages to the third phalanges of the second and third fingers of each hand. It was evident from preliminary explorations that in this position, rather than in the center of the palms, the contact pressure (and hence the amount of current permitted to flow) was relatively constant.

Electric shocks were delivered from a coil built into

an electronic timing mechanism. The intensity was varied with an ordinary tuning dial graduated in arbitrary units. The shock was of 75 msec. duration and was delivered across the *S*'s wrists through two polished brass electrodes.

Conditioned and other tonal stimuli were produced by a type H.1. Advance audiofrequency generator, delivered to *S* through an Erikson low impedance earphone to the preferred ear, the headset being balanced by a nonfunctional earpiece. The frequencies employed were 1967, 1000 and 468 c.p.s. Since the work of Hovland (10) has provoked considerable discussion, three of his frequencies were employed, but no check was made upon the accuracy of his psychophysical measurements. All tonal stimuli were presented at an intensity of approximately 40 decibels, but were equated for loudness in preliminary studies.

A specially built electronic device served to control the presentation of all stimuli at the touch of a post office relay switch. Tonal stimuli lasted for exactly 400 msec., followed during conditioning by an interval of 100 msec., after which the shock of 75 msec. duration was presented.

### *Pretesting Procedure*

In all three groups, *S* was seated in one cubicle, while *E*, after installing and instructing him, controlled the whole conditioning procedure remotely from a separate cubicle. Each *S* was seated in a reclining chair with a footrest. He was first asked to relax for about three to five minutes during which period the basal resistance level typically assumed some steady point. The resistor $R_2$ in the galvanic circuit was then adjusted until a predetermined current of 59.3 microamps was passing through him, thus permitting measurement of basal resistance level to the nearest 1000 ohms. This initial adjustment, of course, placed a certain potential across *S*, which remained constant for him during the entire experimental session.

Pretesting of stimuli involved presentation of the 1967 c.p.s. tone once per minute for three minutes, the GSR being observed on each occasion. It was evident in the vast majority of cases that adaptation was then complete, since, at the fourth pretesting presentation of the tone, the response was either zero or negligible. No pretests of responsiveness to the other two tones were made, it being assumed that generalization of adaptation would be considerable if not complete. Finally, each *S* was given three shocks of increasing intensity and was asked on each occasion for a verbal report of how each felt (unpleasant but not painful, painful, etc.). Successful attempts were made to deliver the same mean intensity of shock to each group of *S*s.

### *Conditioning Procedure*

After a pause of five minutes, each *S* was given eight reinforced trials of the 1967 c.p.s. tone (CS) followed by the shock (UCS) at a staggered rate of one per minute. For each stimulation, minimal and maximal galvanometric readings were taken. One minute after the eighth reinforcement trial, each *S* was subjected to experimental extinction over eight trials. For the latter purpose, one-third of each group (i.e., normal control, anxiety, and schizophrenic groups) was presented with the 1967 c.p.s. tone (*0*), another third with the 1000 c.p.s. tone (*1*), and the remainder with the 468 c.p.s. tone (*2*). Initial and maximal galvanometric deflections were noted through the eighth extinction trial.

### *Treatment of Data*

Haggard (3) demonstrated that logarithm of conductance change is a favorable transformation of GSR data, since it has been found to satisfy the requirements of normality and independence of response magnitude from basal resistance level. Initial examination of data at the first acquisition trial indicated that this measure was superior to both resistance change and conductance change according to these criteria.

## RESULTS

As a matter of course, the basal resistance levels at the beginning of the experiments were compared on an interpathological basis. Mean values in thousands of ohms and variabilities are shown in Table 1. Using separate estimates of the population variance for each group of observations, the appropriate $t$ values showed that the three means differed from each other at less than the .001 level of significance. Further, the variance of the anxiety group was smaller than that of the normal group ($F = 1.58$, $p < .05$), while that of the schizophrenic group was considerably *larger* than that of the normal group ($F = 3.56$, $p < .001$). These figures point up the relatively high variability of schizophrenic patients and, indirectly, the instability of psycho-physiological systems established in such *S*s and reported by others (for example, 9).

A chi-square test of normality of response magnitudes at the first acquisition trial showed that with log conductance transformations the anxiety and normal groups could be considered as showing only random departures from normality (with 5 *df*, $\chi^2 = 8.21$ and $p = .15$ for the anxiety group; with 4 *df*, $\chi^2 = 6.77$ and $p = .15$ for the normal group), while with 5 *df*, the schizophrenic group showed a $\chi^2$ of 12.5. The latter figure would be expected

TABLE 1

BASAL RESISTANCE LEVELS BEFORE CONDITIONING IN OHMS × $10^3$

| | Group | | |
|---|---|---|---|
| | Anxiety | Normal | Schizophrenic |
| Mean | 21.6 | 34.1 | 53.0 |
| *SD* | 3.6 | 5.7 | 20.3 |

to occur only 3 per cent of the time in repeated sampling and hence indicates a significant departure from normality. The first acquisition trial data were also subjected to the $L_1$ test for equality of variance described by Johnson (12), and a value of 0.978 was obtained. Entering Nayer's Table with $k = 3$ and $n = 60$, it is found that a value of $L_1$ equal to or less than 0.967 justifies rejection of the hypothesis of equal variance ($p = .05$). The three groups studied may thus be considered to have initially equal variance.

Product moment correlations between response magnitude and initial level were computed for each group at the first acquisition trial. Values of $r$ were from .120 to −.239. None of these values is significant at the five per cent level. Thus, while there are considerable differences among the groups studied, response magnitude may be regarded as independent of the initial level from which it commenced. A final check on the approximate comparability of the anxiety, normal, and schizophrenic groups was made via analysis of variance of response magnitudes at the first acquisition trial. Since the value of $F$ was but 0.14, it may confidently be concluded that all groups were responding initially with comparable magnitude.

Consequently, rates of extinction were studied via a two-way classification analysis of variance for repeated measurements. Table 2 shows mean response magnitude for each group by extinction trials. Each mean value was derived from the responses of the 60 *Ss* in the group, each third of whom was partly extinguished to one of the three tones employed in the study. In other words, each mean value is actually a regression constant, $a$. It will be observed from Table 2 that there are fairly consistent trends showing the anxiety group to be responding with the greatest, the schizophrenic group with the smallest, and the normal group with intermediate magnitude. The analysis is shown in Table 3. The between groups (i.e., anxiety, normal, and schizophrenic) $F$ ratio of 8.16 ($p < .001$) was obtained using the variance between *Ss* in the same group as the error term. The between extinction trials $F$ ratio of 235.21 ($p < .001$) and the Groups × Trials interaction $F$ ratio of 7.98 ($p < .001$) were obtained using as error term the interaction for Pooled *Ss* × Trials. These results indicate at highly acceptable levels of significance that by the eighth extinction trial all groups had shown a decrease in response magnitude. Further, the magnitude of response was consistently greatest in the anxiety group and smallest in the schizophrenic group, and the absolute intergroup differences in response magnitude were clearly an increasing function of the number of extinction trials. The data are thus in accordance with the original assertion that the anxiety group has the higher and the schizophrenic group the lower total drive level (Hull's $D$). They are in general accordance also with the relevant aspects of recent studies of anxiety and the extinction of conditioned responses (2, 22).

Since heterogeneity of variance among extinction data was not, however, precluded by application of Bartlett's test, thus leaving the acceptability of these results somewhat equivocal, a further test was performed with a nonparametric method devised by Wilson (28). This test has the advantage of being able to explore interaction as well as main effects without any supposition concerning the vari-

TABLE 2

GSR MAGNITUDES[a] DURING EXTINCTION TRIALS (Log. conductance change + K)

| Extinction Trial Number | Group | | |
|---|---|---|---|
| | Anxiety | Normal | Schizophrenic |
| 1 | 1.74 | 1.68 | 1.64 |
| 2 | 1.58 | 1.56 | 1.41 |
| 3 | 1.31 | 1.38 | 1.20 |
| 4 | 1.28 | 1.12 | 1.00 |
| 5 | 1.21 | 0.70 | 0.75 |
| 8 | 0.81 | 0.25 | 0.09 |

[a] Each value in the table is a rounded mean of 60 *Ss*.

TABLE 3

ANALYSIS OF VARIANCE OF EXTINCTION DATA

| Source of Variation | *df* | Mean Square | *F* | *p* |
|---|---|---|---|---|
| Between groups | 2 | 8.770 | 8.16 | <.001 |
| Between *Ss* in same group | 177 | 1.074 | | |
| Between trials | 5 | 39.752 | 235.21 | <.001 |
| Interaction: Trials X Groups | 10 | 1.349 | 7.98 | <.001 |
| Interaction: Pooled *Ss* X Trials | 885 | 0.169 | | |
| Total | 1079 | | | |

TABLE 4
Contingency Table for Extinction Data*

| Trial | Group | | | |
|---|---|---|---|---|
| | Anxiety | Normal | Schizophrenic | $_{b}f_{j}$ |
| 1 | 7 | 9 | 7 | 23 |
| 2 | 13 | 14 | 22 | 49 |
| 3 | 30 | 20 | 25 | 75 |
| 4 | 30 | 39 | 32 | 101 |
| 5 | 32 | 54 | 46 | 132 |
| 8 | 41 | 59 | 60 | 160 |
| $_{b}f_{.j}$ | 153 | 195 | 192 | $540 = n_{b}$ |

* Subscripts $a$ and $b$ indicate that the referent lies above or below the median, and $n_{a} = n_{b}$, and all $n_{ij}$ are equal.

ability of data. The median for the entire set of data was computed and a $2 \times r \times c$ contingency table (see Table 4) was set up, where $r = 6$ (Trials) and $c = 3$ (Groups), and the third factor corresponds to an exact division of the data into two sets, above and below the median. A total value of chi square was computed which was then broken down into main and interaction effects. The resultant analysis (Table 5) yielded a significant between groups value of 12.078 (with 2 *df*, $p < .01$), a significant between trials value of 289.96 (with 5 *df*, $p < .001$), and a significant interaction value of 29.018 (with 10 *df*, $p < .005$). While these results are *statistically* in general accordance with those obtained from the parametric method, there is one important consideration which leaves *interpretation* yet equivocal. Reference to Table 4 shows that when contingency data are used as an indirect measure of response strength during extinction, the normal and schizophrenic populations show virtually zero differences in relation to each other. The significant between groups and interaction values of chi square are thus probably due largely to the predicted deviation of the anxiety group from the other two. On the one hand, such an outcome is possible because of the additive properties of chi square; and on the other, it should be noted that considering the behavior measure employed in this study, resort to a nonparametric method potentially involves the loss of considerable information (cf. the relatively consistent smaller mean response magnitude in the schizophrenic group in Table 2).

Some further light was thrown upon the dubious normal-schizophrenic differences by a comparison of the 60 psychotic *S*s with 36 who

TABLE 5
Summary of Nonparametric Analysis of Extinction Data

| Source | $\chi^2$ | *df* | *p* |
|---|---|---|---|
| Columns (Groups) | 12.078 | 2 | <.01 |
| Rows (Trials) | 289.960 | 5 | <.001 |
| Interaction G × T | 29.018 | 10 | <.005 |
| Total | 331.056 | 17 | |

were not conditionable because of zero GSR to shock, complete GSR adaptation to shock during the acquisition trials, or failure to give any response to the first and subsequent extinction trials. The unconditionable group had been given stronger shock ($p < .01$), showed a much higher resistance level ($p < .01$), were an average of 4.5 years older ($p = .05$), and had been ill for an average of seven years longer ($p < .01$). While certainly suggesting that the schizophrenic population was not randomly sampled, these facts enhance the significance of other results reported here.

## Discussion

Theoretical considerations led at the start of this work to a prediction of strongest conditioning in the anxiety group and weakest conditioning in the schizophrenic group. While the anxiety *S*s clearly showed significantly greater response magnitude during extinction compared with all other *S*s, it was not possible to demonstrate conclusively in any rigorous, formal fashion that the chronic schizophrenic *S*s showed significantly *smaller* response magnitudes than the normal *S*s. But there can be no doubt whatsoever that a significant degree of extinction occurred by the eighth extinction trial, that a significant interaction exists between groups and trials, and that the schizophrenic *S*s show a significantly lower response magnitude than the *anxiety S*s. These general conclusions are thus in accordance with recent studies of questionnaire-selected high and low anxious *S*s (e.g., 2, 22, 24).

At the same time, other recent predictions and conclusions (23, 25), such as the contention that the psychotic conditions autonomically more readily than both the normal *and* the neurotic, are contradicted by the results presented here. The prediction made for the behavior of a schizophrenic in a nociceptive

conditioning situation will, however, more likely be a function of the degree of chronicity and perhaps the deteriorative character of the psychosis. Some of the psychotics studied by Spence and Taylor (23, 25) were outpatients; those here reported upon were all chronic, hospitalized cases, which may partly account for the obtained differences. But the comparison reported above of conditionable and unconditionable schizophrenic *S*s, showing a greater chronicity in the latter group, most strongly suggests a relationship between chronicity and autonomic conditionability.

In the light of the foregoing considerations, it is very difficult to make an overall statement about anxiety, psychosis, and Hull's *D*, unless the relative strengths of *D* can be roughly specified at different stages of the disorder. In other words, some psychotics *may* condition more readily than normals and even than anxious *S*s; but, a priori, the converse is just as likely, depending upon the various qualifications raised here.

## SUMMARY

Conditioning of the GSR to electric shock in 60 hospitalized anxiety states, 60 normal controls, and 60 chronic functional schizophrenics yielded the following conclusions:

1. With magnitude of response (in log conductance units) during experimental extinction as an indirect measure of strength of conditioning, the anxiety *S*s showed significantly stronger conditioning than both the normal and schizophrenic *S*s, the difference in response magnitude being some increasing function of the number of extinction trials, as shown by a significant interaction term. While schizophrenic *S*s showed the lowest *magnitude* of response, it was not possible to offer proof with either a parametric or a nonparametric method of analysis that this magnitude was statistically less than that for the normal *S*s.

2. It was argued that predictions concerning the autonomic conditionability of the schizophrenic must vary with the degree of chronicity of and perhaps the amount of deterioration in the disorder, with less reactivity (for example, Hull's drive strength, *D*) being more typical of the chronic and deteriorated case. A comparison of schizophrenic *S*s who were conditionable with some who were not gave support to such an hypothesis.

## REFERENCES

1. BENDER, LAURETTA, & SCHILDER, P. Unconditioned and conditioned reactions to pain in schizophrenia. *Amer. J. Psychiat.*, 1930, **87**, 365–384.
2. BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
3. HAGGARD, E. A. On the application of analysis of variance to GSR data: III. Some effects of the use of inappropriate measures. *J. exp. Psychol.*, 1949, **39**, 861–867.
4. HILGARD, E. R., JONES, L. V., & KAPLAN, S. F. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–100.
5. HILL, D., ROE, P., THEOBALD, J., & WADDELL, M. A central homeostatic mechanism in schizophrenia. *J. ment. Sci.*, 1951, **87**, 111–131.
6. HOAGLAND, H. Adventures in biological engineering. *Sci.*, 1944, **100**, 63–67.
7. HOAGLAND, H. Enzyme kinetics and the dynamics of behavior. *J. comp. physiol. Psychol.*, 1947, **40**, 107–128.
8. HOAGLAND, H., & PINCUS, S. Steroid excretion and stress of flying. *J. aviat. Med.*, 1943, **14**, 173–193.
9. HOSKINS, R. G. *The biology of schizophrenia*. London: Chapman, 1946.
10. HOVLAND, C. I. Generalization of conditioned responses: II. Sensory generalization of CRs with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.
11. HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
12. JOHNSON, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.
13. LACEY, O. L., & SIEGEL, P. S. An improved potentiometric circuit for measuring the GSR. *Amer. J. Psychol.*, 1948, **61**, 272–274.
14. MALMO, R. B., SHAGASS, C., & DAVIS, J. F. Electromyographic studies of muscular tension in psychiatric patients under stress. *J. clin. exp. Psychopath.*, 1951, **12**, 45–66.
15. MAYS, L. L. Studies of catatonia: V. Perseverational tendencies in catatonic patients. *Psychiat. Quart.*, 1934, **8**, 728–735.
16. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
17. MILLER, N. E. Comments on theoretical models illustrated by the development of a theory of conflict. *J. Pers.*, 1951, **20**, 82–100.
18. PAINTAL, A. S. A comparison of the GSR in normals and psychotics. *J. exp. Psychol.*, 1951, **41**, 425–428.
19. PFAFFMAN, C., & SCHLOSBERG, H. The conditioned knee jerk in psychotic and normal individuals. *J. Psychol.*, 1936, **1**, 201–205.
20. ROSENBAUM, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.
21. SHIPLEY, W. C. Studies of catatonia. VI. Further

investigation of the perseverational tendency. *Psychiat. Quart.*, 1934, **8**, 736–744.

22. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
23. SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.
24. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
25. TAYLOR, JANET A., & SPENCE, K. W. Conditioning level in the behavior disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 497–503.
26. WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, **23**, 83–91.
27. WELCH, L., & KUBIS, J. Conditioned PGR (Psychogalvanic Response) in states of pathological anxiety. *J. nerv. ment. Dis.*, 1947, **105**, 372–381.
28. WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

Received October 31, 1956.

# PARENTAL EVALUATIONS OF CHILDREN AND CHILDREN'S SELF-EVALUATIONS

MALCOLM M. HELPER

*Nebraska Psychiatric Institute, University of Nebraska College of Medicine*

SELF-EVALUATIVE attitudes are widely believed to arise rather directly from evaluations made of the individual by others, with especial influence frequently being attributed to evaluations made by parents (1, 9). In spite of this apparent theoretical agreement, or perhaps because of it, very little empirical study has been made of the relationships between parental evaluations of the child and the child's self-evaluations. Jourard and Remy (6) have made what appears to be the only study having direct bearing on this question. They studied relationships between college students' satisfaction with self and body (self- and body-cathexis) and parental evaluations of the students, the latter being reported by the students themselves. The correlations ranged from .56 to .70 between self-satisfaction expressed by the students and parental satisfaction with the students as reported by the students. These authors note, however, that parental evaluations should be assessed directly before definite conclusions are drawn, because reports by the students as to their parents' attitudes may be subject to systematic distortion.

In the present study, independent assessment was made of self-evaluation in a group of 8th and 9th grade children and of evaluation of the children by their parents.

## METHOD

Forty-two adjectival rating items were developed on the basis of dimensions of personality description found by Cattell to be representative of a very large number of trait names (2). Four scales were also selected from Osgood's analysis of connotative meaning (7). It was believed that these bases for selection would insure a wide sampling of behavior descriptions and avoid undue emphasis on any one aspect of personality description. Each scale was defined at the two extremes by adjectives or adjectival phrases and offered a seven-point scale for rating, as in the example below:

talkative:__:__:__:__:__:__:__:quiet.

Subjects were asked to rate several concepts on each of the forty-six scales by checking the space that best described that concept as they viewed it. Each child was asked to rate himself as he ordinarily thought of himself (actual self-concept), and to rate himself as he would most like to be (ideal self-concept). Each parent used the scale to rate his child as he ordinarily thought of the child (actual child-concept) and to rate his child as he (the parent) most wanted the child to be (ideal child-concept). Other ratings were collected at the same time, and were utilized in investigating other aspects of children's self-description (5).

### *Subjects*

Data were gathered on the 74 children in the 8th and 9th grades in the secondary school operated by the University of Illinois. One or both of the parents of 53 of the children cooperated in providing the ratings described above. Because of incomplete ratings on the part of a few *S*s, the number of families included in different computations varies from 50 to 53. Each computation included all of the families providing pertinent data. Since over one-half of the fathers involved were employed in an academic capacity by a university, the families must be considered to be primarily representative of a narrow portion of the socioeconomic spectrum.

### *Measures of Evaluative Attitudes*

Two methods of measuring attitudes toward (or "acceptance" of) the self and others find current usage. In one method, items are selected that unequivocally indicate a favorable or unfavorable attitude toward the subject in question (3, 8). Ratings on these items can be summed into a single score since the favorability indicated by a given response has been determined in advance. This method has the advantage of simplicity of scoring but is limited to items that have unambiguous evaluational implications. In the second and more popular method, "actual" ratings on a group of items are subtracted from "ideal" ratings on the same group of items. The differences are summed across items to yield a score that reflects the extent of the discrepancy between the actual and ideal concepts. This method (of which the Q sort is a variant) allows for differences among *S*s in evaluating the desirability of a given trait. However, the difference scores are necessarily less reliable than the raw ratings from which they are derived (4).

In the present study, both types of measures were applied to children's self-ratings and to the parents' ratings of their children. A Favorability score was computed by summing "actual" ratings on 15 items on which there was very high agreement among raters as to desirability. These were items on which the ideal self-ratings were heavily concentrated toward one end of the rating scale. All items having a mean ideal self-rating more than 1.5 scale points from the midpoint of the 7-point scale were identified. Of these, the fifteen items having the smallest standard deviation in ideal self-ratings were selected for the Favorability scale. In practical terms, the distribution of ideal ratings on

these items is so narrow that subtracting the actual ratings from them would add little information. The Favorability score was simply the sum of the raw ratings on these fifteen items after all items had been aligned so that a high rating indicated a desirable standing. A Child-Favorability score was obtained from each parent's actual child ratings, while a Self-Favorability score was obtained from each child's actual self-ratings.

A Self-Acceptance score was derived for each child from the actual-ideal discrepancies generated by the remaining 31 items. For each child, the 31 discrepancy scores were squared and then summed in accordance with the specifications for D, the statistic of profile comparison (4). The discrepancies on these items between each parent's actual-child and ideal-child ratings were treated in like fashion to yield a Child-Acceptance score. In this method of scoring, the higher the score, the less the Acceptance. Neither the Self-Acceptance nor the Child-Acceptance scores were distributed normally. For this reason, nonparametric statistical procedures were used in all analyses.

An estimate of each child's acceptability to his peers was available from a sociometric study made by the two classroom teachers involved. Each child had been asked to list classmates with whom he would like to work on a study project. The number of choices received by each child was taken as the index.

The data described above permit study of several questions relating to the problem of self-evaluation. For methodological purposes, the degree of correspondence between the two approaches to the assessment of self-evaluation was studied. The question of primary interest—the degree of correlation between parental evaluations of children and children's self evaluations—was investigated in terms of both the Favorability and the Acceptance variables. The relationship between children's self-evaluations and acceptability to peers was also considered.

## RESULTS

Reliability coefficients were computed by the split-half method for Favorability and for Acceptance in both parents and children. Rank-difference correlation was used because of the non-normality of the Acceptance scores. These coefficients, after correction for total length by the Spearman-Brown formula, are presented in Table 1. The reliabilities are seen to be only moderate, but do appear high enough to justify further study of the measures.

The correlations between the two measures of attitudes toward the child, shown in Table 2, are much lower than their respective reliabilities. The two measures do not appear to tap the same aspects of evaluative attitudes toward the child, either in the parents or within the child himself.

Rank difference correlations between parental Favorability and Acceptance scores and the corresponding self-evaluative scores in children are presented in Table 3. In general, the parent-child correlations are low for both the Favorability and the Acceptance variables. Of the 24 correlations computed, 5 reach significance at the .05 level, 4 fall between the .05 and .10 levels, and 15 do not approach statistical significance. All except one of the correlations fell in the predicted positive direction. (It is not claimed, of course, that the 24 correlations are entirely independent.) On the whole, then, the data do seem to point to the existence of a slight but real tendency toward similarity between parents' evaluations of their children and the children's self-evaluations.

While the correlations do not differ enough among themselves to warrant differential interpretations of the various parent-child combinations, it can be noted that all but one of the significant and near-significant correlations are between corresponding variables at the parent and child levels. That is, parental Favorability has significant relationships only with Self-Favorability in the child, while parental Acceptance is related much more

TABLE 1

RELIABILITY OF FAVORABILITY AND ACCEPTANCE SCORES*

| Group | *N* | Favorability | Acceptance |
|---|---|---|---|
| Boys | 34 | .86 | .77 |
| Girls | 40 | .77 | .80 |
| Fathers | 53 | .66 | .80 |
| Mothers | 51 | .85 | .83 |

* Coefficients are rhos corrected by the Spearman-Brown formula for total length of the scales.

TABLE 2

INTERCORRELATIONS OF FAVORABILITY AND ACCEPTANCE MEASURES

| Group | *N* | Rho |
|---|---|---|
| Boys | 34 | .24 |
| Girls | 40 | .21 |
| Fathers | | |
| Of Boys | 20 | .10 |
| Of Girls | 30 | .18 |
| All Fathers | 50 | .16 |
| Mothers | | |
| Of Boys | 21 | .42 |
| Of Girls | 30 | .23 |
| All Mothers | 51 | .39* |

* Significant at the .01 level.

TABLE 3
PARENT-CHILD CORRELATIONS (RHOS) FOR TWO MEASURES OF EVALUATION OF THE CHILD

| | Boys | | | Girls | | | All *Ss* | | |
|---|---|---|---|---|---|---|---|---|---|
| | *N* | Favorability | Acceptance | *N* | Favorability | Acceptance | *N* | Favorability | Acceptance |
| Fathers | | | | | | | | | |
| Favorability | 20 | .44* | .26 | 30 | .26 | .06 | 50 | .32** | .11 |
| Acceptance | 20 | .18 | .42* | 30 | .28 | .44** | 50 | .31** | .37** |
| Mothers | | | | | | | | | |
| Favorability | 21 | .20 | .10 | 30 | .33 | .18 | 51 | .22 | .08 |
| Acceptance | 21 | .29 | .08 | 30 | −.15 | .39** | 51 | .16 | .27* |

* Coefficient reaches .10 level, two-tailed test.
** Coefficient reaches .05 level, two-tailed test.

consistently to children's Self-Acceptance than to children's Self-Favorability.

The small size of the parent-child correlations does not rule out the possibility of more substantial relationships in certain subgroups. In view of the low correlations between parental Acceptance and Favorability, it would seem reasonable to study the possibility that different combinations of the two variables may be associated with differences in children's self-evaluations. One could reason that children whose parents view them unfavorably would be rather highly sensitive to their parents' acceptance of them and their faults, even though the children's Self-Favorability does not reflect the parents' basically negative view of them to any great extent. On the other hand, children whose parents view them as basically sound might well be less sensitive to any unusually high aspirations (high actual-ideal discrepancy) that the parents might express for them. In terms of the present variables, there should be a relatively high relationship between parental Acceptance and children's Self-Acceptance (or Self-Favorability) in cases where parents have low Favorability scores, but not in cases where parents have high Favorability scores.

In accordance with this hypothesis, the relationship between parental Child-Acceptance and the two self-evaluative variables was studied separately in the group of families characterized by high parental Favorability and in the group characterized by low Favorability. The median Favorability score was used as the dividing line. Separate analyses were carried out for mothers and for fathers. In each low Favorability group, children receiving high and low parental acceptance were compared with respect to Self-Acceptance and Self-Favorability. Similar comparisons were made in each high paternal Favorability group. The Mann-Whitney *U* test was used in making all statistical tests. The results of this procedure can be described briefly: The one significant difference occurred in the group characterized by low Favorability of the mother. Here, as expected from the hypothesis, children receiving low Acceptance from the mother had lower Self-Acceptance than did those receiving high Acceptance from the mother. However, the preponderance of nonsignificant differences indicates that, in terms of the present data, the hypothesized effects are slight at best, and probably should be discounted completely without further evidence.

TABLE 4
COMPARISON OF HIGH AND LOW SOCIOMETRIC GROUPS ON SELF-EVALUATIVE VARIABLES
(U-test)

| | *N* | *N*[a] | Self-favorability | | Self-acceptance | |
|---|---|---|---|---|---|---|
| | | | Sum of ranks | *p* | Sum of ranks | *p* |
| Boys | 34 | 15 | 221.5 | — | 160.5 | .001 |
| Girls | 40 | 22 | 423.0 | — | 466.0 | — |

[a] Number of cases in low sociometric group.

To study the relationship between children's self-evaluations and their acceptability to their peers, the distribution of choices received in the sociometric procedure was first dichotomized for each sex separately. For each sex, the *U* test was employed to evaluate the difference in Self-Favorability and Self-Acceptance between the more and the less popular children. These results are presented in Table 4. The popular boys express a significantly higher degree of Self-Acceptance than the

unpopular boys. No significant differences emerge for girls.

## DISCUSSION

In interpreting the rather tenuous relationship observed between parents' evaluations of children and the children's self-evaluation, we should first take into account the problem of the validity of the measures used. While the children's verbal responses can, if desired, be taken as operational definitions of their self-evaluative traits, the parental variables specified by the theoretical position mentioned at the outset are those which reflect actual evaluative behavior by the parent toward the child over a span of several years. From this point of view, it is rather encouraging that an apparently significant relationship does emerge when only rating data are available from the parents.

The contrast between the present low correlations and the relatively high correlations reported by Jourard and Remy (6) can be interpreted in two basically different ways. It is possible, as those authors note, that the children's reports of their parents' attitudes are distorted (possibly by response set) in such a way as to coincide spuriously with the children's self-ratings. On the other hand, it is also possible that children's reports are actually more valid as measures of parental evaluations than ratings by the parents themselves. This uncertainty would seem to call for validation of both types of measures against observed parental evaluative behavior.

Regardless of the size of the true correlations between parental attitudes and children's self-evaluations, no causal interpretations are justified until temporal trends in the two variables can be studied. This type of study would represent a logical step from the presently available evidence of a positive relationship between parental evaluations of the child and the child's self-evaluations. Although psychodynamic interpretations are premature at the present time, it does seem wise, in view of the small size of the correlations found, to note that parental evaluations may have very great consequences for the child's personality even if they are not reflected directly in the child's self-evaluations. In fact, a study of the reactions of children whose parents evaluate them negatively might well shed much light on defensive processes.

The evidence that Acceptance and Favorability are substantially independent variables is rather difficult to interpret theoretically since the two types of measures were originally distinguished on methodological grounds and were expected to reflect the same attitudinal variables. It was suggested in comparing the effects of the various combinations of parental Favorability and Acceptance that Favorability reflects simply a description of an individual in evaluative terms while Acceptance indicates the attitude toward the individual with his perceived defects or assets. This differentiation would suggest that Acceptance is the more emotionally loaded of the two variables and has more significance for interpersonal relationships. While the data of the present study do not resolve this question, it can be noted that the single significant relationship between self-description and sociometric status involved the Acceptance variable. While the present results provide only the scantiest basis for interpretation of these two aspects of self-evaluations, they do point up the need for further methodological study of such constructs as self-acceptance and self-esteem.

## SUMMARY

Two self-evaluative measures, labelled Self-Favorability and Self-Acceptance, were derived from self-ratings by 74 8th- and 9th-grade children. Analogous measures of parental Favorability and Acceptance with respect to the child were derived from ratings by parents of 51 of the children. Sociometric information was available for the children. Although the measures of Favorability and Acceptance had adequate reliability, correlations between the two variables were low in both parents and children. Correlations between parental evaluations and children's self-evaluations tended to be small but consistently positive. The results were interpreted as evidence that attempts to increase the validity of assessments of parental attitudes and further to differentiate aspects of self-evaluations would prove rewarding.

## REFERENCES

1. CAMERON, N., & MAGARET, A. *Behavior pathology.* New York: Houghton Mifflin, 1951.
2. CATTELL, R. B. *Personality.* New York: McGraw-Hill, 1950.

3. CRANDALL, V. J., & BELLUGI, U. Some relationships of interpersonal and intrapersonal conceptualizations to personal social adjustments. *J. Pers.*, 1954, **23**, 224–232.
4. CRONBACH, L. J., & GLESER, G. C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.
5. HELPER, M. M. Learning theory and the self concept. *J. abnorm. soc. Psychol.*, 1955, **51**, 184–194.
6. JOURARD, S. M., & REMY, R. M. Perceived parental attitudes, the self, and security. *J. consult. Psychol.*, 1955, **19**, 364–366.
7. OSGOOD, C. E., & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.
8. SOLLEY, C. M., & STAGNER, R. Effects of magnitude of temporal barriers, type of goals and perception of self. *J. exp. Psychol.*, 1956, **51**, 61–70.
9. SNYGG, D., & COMBS, A. W. *Individual behavior*. New York: Harper, 1949.

Received December 19, 1956.

# IDENTIFICATION IN GROUP THERAPY[1, 2]

GERALD J. BRISKIN

*Byron Harless and Associates, Miami, Florida*

THE process of psychotherapy implies changes in behavior, and the variables associated with these changes have become a focal point for research in therapy. In a recent study, Schrier (11) found that identification with the therapist in individual therapy was significantly related to therapeutic success. This suggests that the ability to identify with the therapist is an important variable in the prognostic evaluation of candidates for therapy. The intent of this study is to investigate two of the variables contributing to identification.

For the purposes of this paper, identification is operationally defined as changes in patient behavior from pretherapy status to behavior similar to that manifested by the therapist. The variables selected for study were rigidity and lability.

Rigidity is a well-studied entity. Considerable research indicates that certain individuals have difficulty in changing their patterns of behavior. They seem encysted against the world and behave in a stereotyped and inflexible manner. Tension and anxiety are well "bound" in this personality configuration, and emotional content is effectively removed from thoughts and ideas to the extent that anxiety is not strongly experienced by this kind of individual in therapy. Fenichel (5) suggests that identifications are always motivated by drives or tension. Since the rigid individual may not experience tension very readily, and since access to and from the ego is limited, it is quite likely that he will have difficulty in identifying.

The term lability, used in this context, refers to a personality pattern that closely approximates the preschizophrenic personality in which the ego boundaries are vaguely defined and weakly maintained. An individual possessing this kind of ego structure has difficulty in distinguishing between self and non-self. Because of this, his reality-testing abilities are limited, and he often confuses events occurring in the external world with events occurring in himself. As a result, his ability to identify is severely impaired.

Based upon this reasoning, two hypotheses were formulated: (*a*) There is an inverse relationship between rigidity and identification; (*b*) there is an inverse relationship between lability and identification.

## METHOD

The *S*s utilized in this study were ninth- and tenth-grade male and female high school students. They were selected for the experiment by asking members of the social living classes if they would like to volunteer to participate in groups that would discuss personal problems of adolescents. They were told that only a limited number could be selected and that they would be excused from class to participate. A total of 109 students volunteered. All of the volunteers were then rated by their social living class teachers on a form designed to assess emotional adjustment. Those students who were rated most poorly adjusted were then eliminated from the study on the ground that they would be least likely to benefit from short term therapy.

Three groups were selected to participate in therapy, each composed of five boys and five girls. Before the therapeutic groups were formed, *S*s were asked to fill out a specially designed questionnaire,[3] and were given group forms of the Rorschach. These forms and tests were administered by members of the research staff other than the therapists.

Following this collection of data, the groups started in therapy, each group having a different therapist. The groups met twice a week over a period of four months, with interruptions for school holidays. *S*s were told that they were free to drop out of the groups up to the fifth session, but they were expected to stay in the group after the fifth meeting. There was no other pressure to compel attendance.

The therapeutic approach followed group analytic principles, with the therapist very active in serving as the focal point of the group, yet being essentially per-

[1] Based upon a Ph.D. dissertation written at the University of Michigan. The author wishes to express his appreciation to Max L. Hutt, Donald Marquis, Theodore Newcomb, Edward Bordin, and Alvin Zander, who contributed much to the thesis. Gratitude is also due the Rackham Fund, whose grant supported the Group Therapy Project on which this work is based.

[2] Paper delivered at the APA, September, 1954.

[3] To save printing costs, a 3-page tabular presentation of the questionnaire has been deposited with the American Documentation Institute. Order Document No. 5465 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

missive and accepting of all behavior, including fairly extreme forms of acting out. This technique is based on Wender's (13) notions of group therapy and is designed to elucidate and clarify defensive behavior and its significance.

At the conclusion of therapy, all *S*s again filled out the same questionnaire that they had at the beginning of the study. These questionnaires were administered by the therapists.

Each of the therapists was rated on this same questionnaire by experienced third- or fourth-year clinical psychology graduate students who knew them well but who were not involved in any competitive relationships with them. These judges were instructed to answer the questionnaire as they believed the therapist would. A combined rating for each therapist was then obtained by averaging the judges' ratings.

Having the therapists rated by judges seemed a sounder procedure than having the therapist rate himself or having him rated by his group members. With regard to the first alternative, the therapist, regardless of his insight, may still tend to rate himself defensively and thus present a distorted picture of himself. Cutler (4) made use of this reasoning in utilizing discrepancies between therapists' self-ratings and observer ratings as an index of therapists' areas of conflict. The other alternative, on casual examination, seems to have considerable merit. However, since *S*s in therapy may be involved in all sorts of transference relationships with the therapist, any ratings of him by them may also be distorted. The use of external observers was therefore decided upon.

The questionnaire defining identification in therapy derived from Murray's (8) inventory was altered to fit the needs of an adolescent group. It consisted of 50 statements, each of which could be answered *always, at times, seldom, never*. Each item was a description of behavior or feelings—for example: "I feel I should like to be a leader of a group," or "I try to be good in home and in school."

The Rorschach test was essentially a group adaptation of the individual administration. The Harrower-Erickson Rorschach slides were used. The protocols were scored for Beck's (1) locations, content, form levels, and populars and for Klopfer's (7) determinants. Since group Rorschachs differ considerably from individual Rorschachs, especially as to the inquiry, it was necessary to adhere to certain additional rules: If a percept was named in the free association but there was no inquiry or location, it was not scored. If a percept was named in free association and there was a location but no inquiry, it was scored F only for determinants. Precision alternates were scored only once. Otherwise, the records were scored in accordance with the usual rules specified by Klopfer and Beck. No reliability study of the scoring was done since numerous reports in the literature (10, 12) indicate that high interjudge reliability can be expected between Rorschach scorers of considerable experience.

The measure of identification utilized in this study was operationally defined as the total number of shifts from each *S*'s pretherapy questionnaire to his posttherapy questionnaire in the direction of the therapist as rated by the observers. The measures of lability and rigidity were derived from the Rorschachs. However, since the determinants obtained from a group administration of the Rorschach vary considerably from those obtained in individual administration, it was not possible to apply standard interpretive procedures to the Rorschach records without first evaluating the data. In order to accomplish this evaluation, the Rorschachs for the volunteers who did *not* receive therapy were scored and distributions made of all determinants. Since most of these distributions were greatly skewed, medians and quartile deviations were calculated rather than means and standard deviations. These statistics were then used as standards in setting up the rating scales.

However, Rorschach variables do not scale proportionately; a total of four M responses, for example, is not twice as significant as two M responses. Therefore, in assigning weights to Rorschach variables, cutting points must be determined in accordance with the shape of the distribution of the variables dealt with. Furthermore, since most Rorschach work has been done with individually administered tests, the "accepted" cutting points must also be modified in terms of the group administered distributions. An example follows:

> It is a standard Rorschach procedure to regard a protocol in which F comprises more than fifty per cent of the responses as constricted. But the median F percentage for the group records was 74.5. Two explanations are available: The population dealt with produced unusually frequent F percepts, or the lack of an individually conducted inquiry resulted in many responses being scored F which otherwise would not have been so scored. Since the latter explanation seems more tenable, the notion of the cutting point for constriction as 50 per cent was discarded, and the cutting point used was based on Q3, which was 88. Thus, a weighted scale for constriction could be established with scores of 87 per cent or below weighted at zero, 88–95 per cent at one, and 96 per cent or over at two.

Based on this procedure, the rating scales for lability and rigidity were developed. To determine the rigidity or lability scores for a particular *S*, the determinants from his Rorschach were given the appropriate factor weights as indicated in the respective scales, and these weights were then totaled. Since some negative scores were obtained, all of the scores had a constant added to them so that positive numbers could be dealt with.

The measure of rigidity was based upon pertinent theoretical and clinical positions as found in the literature (2, 6, 7, 10). One study of particular relevance by Cowen and Thompson (3) related certain Rorschach findings to measures of Einstellung rigidity. The rigidity index may be found in Table 1.

As discussed previously, the measure of lability is closely related to what has been described as the preschizophrenic personality. Consequently, the scale measuring this configuration was derived from standard Rorschach "signs" (2, 6, 7) considered as indicative of such a personality. These signs and their weights are given in Table 2.

The interobserver agreement of ratings of the therapist can be found in Table 3. These agreements are given in terms of coefficients of contingency rather than the usual product-moment correlations. This statistical treatment was necessitated by the nature of

TABLE 1
THE RIGIDITY RATING SCALE

| Factor | Weight |
|---|---|
| I. Total Responses | |
| 11 or less | 3 |
| 27 or more | 3 |
| II. D plus Dd% | |
| 85 or more | 3 |
| 25 or less | 3 |
| III. Sum M | |
| 0 | 3 |
| 1 | 0 |
| 2 or more | −3 |
| IV. F% | |
| 85 or more | 3 |
| 50 or less | −3 |
| V. F + % | |
| 90 or more | 3 |
| 60 or less | −3 |
| VI. A% | |
| 60 or more | 3 |
| 30 or less | −3 |
| VII. Number of Content Categories | |
| 4 or less | 3 |
| 9 or more | −3 |
| VIII. Populars | |
| 6 or more | 3 |

TABLE 2
THE LABILITY RATING SCALE

| Factor | Weight |
|---|---|
| I. Sum CF plus C greater than FC (FC = .5, CF = 1.0, C = 1.5) | |
| by 0–.5 | 0 |
| by 1.0–2.0 | 2 |
| by more than 2.0 | 4 |
| II. Total M | |
| 0 | 3 |
| 1 | 0 |
| 2 or more | −3 |
| III. F% plus F + % | |
| F% greater than 80, F + % less than 65 | 6 |
| IV. Total FM | |
| 0–1 | 0 |
| 2 or more | 3 |
| V. M quality | |
| For each flexor | −2 |
| For each extensor | 2 |
| VI. Pure H% (does not include human like percepts) | |
| 15 or less | 3 |
| 30 or more | −3 |
| VII. Card Rejections | |
| 0–1 | 0 |
| 2 or more | 3 |

the data. Ratings on the questionnaire could be made only in terms of four categories, *always*, *at times*, *seldom*, *never*. Since the distributions of these ratings were skewed, and since product-moment correlations are attenuated by a limited range (only four categories in this sample), the usual correlational techniques were inappropriate and, consequently, the contingency coefficient was substituted. A fourfold table was used, grouping *always* and *at times* in one category and *sometimes* and *never* in the other.

TABLE 3
COEFFICIENTS OF CONTINGENCY*** FOR INTERJUDGE AGREEMENT IN RATING THERAPISTS

| Therapist A | Therapist B | Therapist C |
|---|---|---|
| .36** | .40** | .36* |

* .05 level of confidence.
** .01 level of confidence.
*** Upper limit of coefficient of contigency = .7071.

## RESULTS

The hypothesis specifying the relationship between identification and rigidity was tested by correlating the identification scores of *S*s in the therapy groups with their rigidity scores. A product-moment correlation of −.73 was obtained, significant at the .01 level. In order to test the possibility that rigid individuals might show less identification with the therapist simply by making fewer shifts on the questionnaire, regardless of direction, as nonrigid *S*s, the total number of shifts for each *S* was correlated with his respective rigidity score. The product-moment correlation was .09. A $t$ test between the differences of the mean number of shifts for the rigid and nonrigid group was similarly not significant. The possible objection therefore seems irrelevant, and the findings may be interpreted as supporting the hypothesis.

The hypothesis specifying the relationship between identification with the therapist and lability was also tested by correlating the identification score of each individual in therapy with his lability score. A product-moment correlation of −.64 was obtained, significant at the .01 level. This finding is also interpreted as supporting the hypothesis.

## DISCUSSION

The primary intent of this study was to make an exploratory attempt to evaluate two hypotheses derived from psychoanalytic theory as they apply to group therapy. In general, both hypotheses were supported. These findings lend emphasis to the role that ego structure plays in identification and to the fruitfulness of research on ego functions.

Previous studies have suggested that the ability of patients to identify with the thera-

pist markedly increases the probability of therapeutic success. Since there are far more prospective patients than the limited number of psychotherapists can accommodate, the selection of candidates for therapy is generally made on the basis of an estimated ability to benefit from treatment. In this respect, the present findings imply potentially useful criteria for the prediction of identification, which may be a highly relevant factor in prognostic evaluation.

## SUMMARY

The *S*s of this study were male and female adolescents (12–15 years of age) who were asked to volunteer to participate in groups that "would discuss adolescent problems." Those who volunteered were rated by their teachers on emotional adjustment. *S*s rated as most poorly adjusted were eliminated from the study. Three therapy groups were organized, each composed of five boys and five girls, each in the charge of a different therapist. The members of the group were given a special questionnaire, designed to assess needs, and a group form of the Rorschach before therapy began. At the conclusion of therapy, all *S*s again took the questionnaire. Two or more judges rated each therapist on the questionnaire items.

A measure of identification was obtained by taking the number of shifts by the *S*s from pre- to posttherapy questionnaires which were in the direction of the therapist as defined by the combined observer ratings. Measures of lability and rigidity were derived from the Rorschach tests.

The hypotheses that both rigidity and lability were inversely related to identification were tested by correlational methods. Results were positive in both instances.

## REFERENCES

1. BECK, S. *Rorschach's test: Basic processes*. Vol. I. New York: Grune & Stratton, 1949.
2. BECK, S. *Rorschach's test: Advances in interpretation*. Vol. III. New York: Grune & Stratton, 1952.
3. COWEN, E. L., & THOMPSON, G. G. Problem solving rigidity and personality structure. *J. abnorm. soc. Psychol.*, 1951, **46**, 165–176.
4. CUTLER, R. The relationship between therapists' personality and certain aspects of psychotherapy. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
5. FENICHEL, O. *The collected papers of Otto Fenichel*. First Series. New York: W. W. Norton, 1953.
6. KLOPFER, B., AINSWORTH, M. D., KLOPFER, W. G., & HOLT, R. R. *Developments in the Rorschach technique*. Vol. I. New York: World Book, 1954.
7. KLOPFER, B., & KELLEY, D. M. *The Rorschach technique*. New York: World Book, 1946.
8. MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
9. RORSCHACH, H. *Psychodiagnostics*. Berne: Verlag Hans Huber, 1949.
10. SCHNEIDER, S. F. Prediction of psychotherapeutic relationship from Rorschach's test. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
11. SCHRIER, H. The significance of identification in short-term therapy. Unpublished doctoral dissertation, Univer. of Michigan, 1951.
12. SILVERMAN, H. The prediction of consciousness of conflict in the self from the Rorschach. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
13. WENDER, L. The dynamics of group psychotherapy and its applications. *J. nerv. ment. Dis.*, 1936, **84**, 54–55.

Received February 1, 1957.

# SOME RELATIONSHIPS AMONG PROJECTIVE TEST, INTERVIEW, AND SOCIOMETRIC MEASURES OF DEPENDENT BEHAVIOR[1]

BERNARD J. FITZGERALD[2]
*VA Hospital, Roanoke, Virginia*

THE need for experimental evaluation of projective test data as used for describing and evaluating behavior has frequently been pointed out (4, 7). Some authors disparage such experimental evaluation, suggesting that perhaps clinically useful evidence is distorted or lost in the process of quantification (2, 3). However, it would seem that systematic investigation of projective test data is essential, particularly in view of the failure of many studies to predict successfully from projective test responses to overt behavior, even on the basis of clinical evaluation of the data.

The present study was developed using descriptive constructs from Rotter's social learning theory (6). From the point of view of social learning theory, any given situation has within it possibilities for the satisfaction of many different needs. Which of these needs an individual will strive to satisfy will depend upon two factors, the *value* that the satisfaction of a given need has for him and his *expectancy* that his behavior will lead to that satisfaction. The value which the satisfaction of a given need has for the individual is termed Need Value. In the measurement of Need Value, emphasis is upon what need satisfactions the individual *prefers*. An individual's expectations that particular behaviors will lead to the desired need satisfaction is termed Freedom of Movement. Emphasis is upon what the individual *expects* to happen. Freedom of Movement is high if the individual expects his behaviors to lead to need satisfaction. An individual's Freedom of Movement is low if he expects failure or punishment for such behaviors. One method of measuring Freedom of Movement depends upon identifying the frequency with which the individual, when seeking any particular need satisfaction, resorts to symbolic rather than direct, overt behavior to obtain that satisfaction; the more he resorts to symbolic behavior, the lower is his Freedom of Movement.

In terms of this study, it was hypothesized that if projective test responses represent, in part, symbolic ways of obtaining gratification, then frequency of response should be highest when Freedom of Movement is low and Need Value high. This hypothesis assumes that a projective test provides a situation in which symbolic gratification is possible; therefore, the individual tends to give responses involving satisfactions that have a high value for him but for which he has a low expectancy of realization. Such a situation, where Need Value is higher than Freedom of Movement, is considered in this study to be a situation of important conflict.

The purpose of this study was to provide data related to the interpretation of needs as reflected in the content of projective test responses. Dependency, the need investigated, may be defined as the need to have other individuals prevent frustration or punishment and provide for the satisfaction of other needs. The projective tests used were the Murray TAT and the Rotter Incomplete Sentences Blank (ISB). The measures of overt behavior were a sociometric technique and an interview. The measures of Need Value and Freedom of Movement for Dependency were also derived from the interview.

The hypotheses of this study are as follows:

1. Frequency of n Dependency responses in a projective test is positively related to sociometric and interview indices of Dependent behavior; and

2. Frequency of n Dependency responses in a projective test is positively related to the importance of conflict in n Dependency.

## METHOD

Eighty *Ss*, 20 from each of four social fraternities, participated in the study. All *Ss* had been members of their respective fraternities for at least six months and students at the Ohio State University for at least one year. A pilot study was devoted to the development of the interview, the sociometric measure, and judges'

[1] This paper is based on a doctoral dissertation done at the Ohio State University. The author acknowledges with gratitude the advice and guidance given him by J. B. Rotter, under whose direction this study was carried out. The writer also is grateful to Gordon Rader for his helpful criticism of this manuscript.

[2] Now at Philadelphia State Hospital.

manuals for scoring interview and projective test responses.

The criteria of overt behavior were a sociometric index and an interview measure of Dependent behavior. A nominating technique was used as the sociometric measure. Two descriptions were provided the *Ss* in order to make their nominations. One of these descriptions was worded to describe certain behaviors considered most and another worded to describe behaviors least representative of Dependent behavior. Care was taken not to label the descriptions as representative of any particular kind of behavior, e.g., Dependent. *Ss* were asked to select from the members of the group those who, in general, best fit one or the other of the two descriptions. The nominations were made alternately, i.e., the one *S* who behaved most Dependently, the one *S* who behaved most non-Dependently, and so on. In this manner, then, all *Ss* made three nominations for each description.[3]

A structured clinical interview was developed to elicit from these same *Ss* historical data relevant to their Dependent behavior and data relevant to their Need Value and Freedom of Movement for Dependency satisfaction. The interview was divided into eleven areas: school, home, advisor, major academic subject, roommate, friends, girl friends, health, personal problems, and religion. A manual for scoring the interview variables of Dependent behavior, Need Value, and Freedom of Movement for Dependency satisfaction was developed. Included in the manual were the definition of Dependency and a set of general principles derived from the definition, as well as examples from interview protocols obtained in the pilot study. Examples were derived from these interviews to correspond to the points 1, 4, 7, and 10 on a ten-point rating scale. The in-between points were included to allow finer discriminations than were possible from the selected referents in the manuals. The values attached to the examples in each of the manuals were arrived at by the experimenter's first assigning crude values consistent with the context from which the examples were taken. Two independent raters also assigned values to these examples, and the scale values were adjusted to maximize agreement among the three raters. Finally, these refined scales were discussed with the interview judges. Changes were made in the scales where necessary for complete agreement among the judges, and the examples were then assigned final values.

Judges' scoring manuals were similarly developed for the TAT (Cards 3BM, 4, 6BM, 7BM, 8BM, 10, 12M, 13MF, 15, 17BM, 18BM, 20) and the ISB. From the TAT stories obtained in the pilot study, a manual was prepared to aid in the determination of the presence or absence of a Dependency plot for any given story. In the development of the manual, each pilot protocol was carefully read, and any story which could be considered to have a Dependency plot was noted. Two projective test judges and the experimenter discussed those stories which were "borderline" or difficult to judge for n Dependency. Following discussion of each borderline story, the three judges agreed on each story as representative of either Dependency or non-Dependency. These stories, as well as stories clearly representative of n Dependency, were included in the final form of the manual. Dependency plots were identified in relation to both the Hero and the Other Figures. Suggestions were included in the manual to promote agreement among the judges in the determination of the Hero figure. From ISB responses obtained in the pilot study, as well as from a large pool of items gathered previously by others, a manual for scoring ISB responses for n Dependency was developed in a similar fashion. In making their ratings, the judges were asked to use the manuals by comparing the TAT or ISB item to be rated with examples included in the manuals. Approximately 30 hours were spent in training the judges in the use of the manuals.

[3] The method of nominating only the extremes of a group has been found to be at least as reliable as rating an entire group and to provide a distribution of ratings similar to that obtained when each member of a group is to be rated by every other member.

### *Procedure*

The *Ss* were first seen at their respective fraternity houses, where the sociometric and projective measures were obtained. Following the group's completion of the sociometric ratings, the TAT cards were projected on a screen. *Ss* were allowed six minutes, the TAT card remaining in view, to write each story. Finally, the ISB was administered. Each *S* was interviewed individually within a two-week period following the group testing. All interviews were recorded on the Gray Audograph, and the judges made their ratings of the interviews from these recordings. The experimenter and two interview judges, working independently, rated the interviews. The experimenter and two different judges, working independently, rated the projective test responses. The protocols and interviews were coded so that the identity of the individual *Ss* was not known to the judges. All TAT and ISB responses for any given card or item were rated before continuing to the next item, and all responses were rated by at least two judges.

Estimates of the interrater reliabilities for the interview variables, the TAT, and the ISB were obtained by correlating (Pearson *r*) the ratings obtained from each judge with those of each of the other two judges. The interview variables were each rated on a ten-point scale. For the TAT and the ISB, each *S* was given a score by counting the number of stories or items rated as Dependent. The average interrater correlations for the interview were as follows: Dependent behavior, .54; Need Value, .45; Freedom of Movement, .56. The average interrater correlations were for the TAT (Hero), .75; TAT (Other), .11; and the ISB, .78. Because of the unreliable TAT Other Figure ratings, this index was dropped from consideration. The sociometric ratings yielded corrected split-half reliability of .83.

## RESULTS

A test of the first hypothesis concerning the relationship between projective test responses and overt Dependent behavior was made by correlating (Pearson *r*) the interview and the sociometric ratings of Dependent behavior with the frequency-of-response ratings on the

TAT and on the ISB. It appears from Table 1 that there is no evidence of a relationship between frequency of response in n Dependency on the TAT and either the interview or the sociometric measures of Dependent behavior. The ISB, in terms of frequency of response in n Dependency, yields indices which are related to both at low but significant magnitudes.

The second hypothesis concerning the relationship of frequency of response in n Dependency to importance of conflict in n Dependency, as determined from the interview, was evaluated by the chi-square test. In order to make this test, *Ss* were divided into two groups for each projective test respectively, i.e., those *Ss* with a high number (four or more on the TAT; seven or more on the ISB) vs. those with a low number (three or less on the TAT; six or less on the ISB) of Dependency responses. The high and low groups for each test were compared on the basis of an importance-of-conflict score which was derived in the following manner: *Ss* with Freedom of Movement equal to or greater than Need Value were considered to have "No Important Conflict" in n Dependency and were compared with those *Ss* whose Freedom of Movement was less than Need Value. These latter *Ss* were considered to have an "Important Conflict" in n Dependency. The results of this comparison are reported in Table 2. It appears that those *Ss* who have a relatively higher Need Value than Freedom of Movement for Dependency satisfaction give significantly more responses in n Dependency on the TAT. The results for the ISB are not significant.

It seems evident also from Table 2 that the TAT tends to elicit dissimilar data, in terms of a frequency-of-response measure, from that elicited by the ISB. Comparisons of these two tests in terms of frequency of response in n Dependency as related to interview and sociometri measures of Dependent behavior (Table 1) support this conclusion. Further evidence that these two instruments tend to elicit dissimilar data is given by their low intercorrelation (.09).

On the other hand, the correlation between ratings of overt Dependent behavior from the interview and from sociometric measures was .59. Important to note here is the potential usefulness of the economical sociometric technique in relation to the laborious clinical interview as a quantified index of certain kinds of overt behavior.

TABLE 1

CORRELATIONS BETWEEN THE MEASURES OF DEPENDENT BEHAVIOR AND FREQUENCY OF RESPONSE RATINGS ON THE PROJECTIVE TESTS

($N$ = 60)

| | TAT | ISB |
|---|---|---|
| Interview | .07 | .28* |
| Sociometric | .06 | .25* |

* $p$ = .05.

TABLE 2

CHI SQUARE OF THE FREQUENCY OF DEPENDENCY RESPONSES ON THE PROJECTIVE MEASURES COMPARED TO CONFLICT IN NEED DEPENDENCY

($N$ = 60)

| Conflict | Total Number of Dependency Responses | | | |
|---|---|---|---|---|
| | TAT | | ISB | |
| | High D Group | Low D Group | High D Group | Low D Group |
| "Important Conflict" (Freedom of Movement less than Need Value) | 88 | 27 | 183 | 79 |
| "Not Important Conflict" (Freedom of Movement equal to or greater than Need Value) | 16 | 18 | 65 | 37 |
| | $\chi^2$ = 12.24; $p$ = .001. | | $\chi^2$ = .99; $p$ = .30. | |

## CONCLUSIONS AND DISCUSSION

The following conclusions, all in terms of frequency of response in n Dependency, may be drawn from this study:

1. There is no evidence that TAT responses are related to interview or sociometric indices of Dependent behavior.

2. TAT responses are significantly related to an interview measure of importance of conflict in n Dependency.

The foregoing conclusions lend support to two assumptions listed by Lindzey in his review of the clinical use of the TAT. One assumption states, "The dispositions and conflicts that may be inferred from the storyteller's creations are *not always* reflected

*directly* in overt behavior." Another holds that, "Themes that are recurrent in a series of stories are particularly apt to mirror the impulses and conflicts of the story-teller" (4, p. 3). The similarity of these assumptions to the hypotheses in this study is evident.

3. ISB responses tend to be related to interview and sociometric measures of overt, Dependent behavior.

4. ISB responses show no evidence of relationship to an interview measure of importance of conflict in n Dependency.

The third conclusion is similar to that of McClelland (5), who, using a Sentence Completion Test to measure n Achievement, argued that this test provides results which are dissimilar to those provided by the TAT and which are more similar to conscious self-judgments. Blyth (1) found also that the ISB differentiated between groups of veterans who behaved dependently or independently in terms of accepting or rejecting offers of psychotherapy.

It is evident that the TAT and the ISB, although both projective tests, provide dissimilar data in terms of frequency of response in a given need category, and that inferences from one test cannot be generalized to the other. More specifically, where inferences about overt behavior are to be made, a highly structured, behaviorally oriented projective test, such as the ISB, may be the more appropriate. The TAT does not appear to be particularly useful as a source of inferences about overt behavior. Where inferences about conflict in a given need area are to be made, a projective test that allows considerable freedom of response, such as the TAT, is probably appropriate. The ISB appears to be less appropriate as a source of inferences about conflict in a given need.

The unmodified TAT and the ISB were used in this study as they generally are employed in clinical psychological work. It is important to emphasize again that the measure used in relation to the projective test responses was a frequency-of-response measure. It is not intended to imply that such a measure of projective test responses is the only possible one or that such a measure is more valid or economical than other possible uses of such projective test responses. In fact, it is evident that a frequency-of-response measure makes no more than minimal use of the data provided by the TAT and the ISB. Nevertheless, with descriptive constructs from Rotter's social learning theory, which provided for an objective and communicable method, a predictive relationship was demonstrated between frequency of response in n Dependency on the TAT and an interview measure of importance of conflict in n Dependency.

## SUMMARY

Data relevant to two hypotheses concerned with the evaluation of projective test responses were obtained in this study. The projective test measure was frequency of response in n Dependency on the TAT and on the ISB. A structured clinical interview and a sociometric technique were used as measures of overt Dependent behavior. The clinical interview also was structured to provide measures of Need Value and Freedom of Movement for n Dependency. Eighty *S*s participated in the study. A pilot study with 20 of these *S*s provided data for judges' scoring manuals for the interview and projective test situations. All *S*s were tested in groups and interviewed individually.

Results of the study may be summarized as follows:

1. Frequency of response in n Dependency on the TAT yielded no evidence of relationship to the interview and sociometric measures of overt Dependent behavior.

2. Frequency of response in n Dependency on the ISB was related to interview and sociometric measures of overt Dependent behavior.

3. Importance of conflict in n Dependency, as derived from the interview, was related to frequency of response in n Dependency on the TAT. The results for the ISB were not significant.

4. Dissimilar data, in terms of frequency of response in n Dependency, were elicited by the TAT and the ISB.

5. A high degree of similarity was found between the sociometric and interview ratings of overt Dependent behavior. This suggests the usefulness of sociometric ratings in lieu of clinical interviews to obtain an approximate, quantifiable measure of certain kinds of overt behavior.

## REFERENCES

1. BLYTH, D. D. Dependency, independency, and other factors related to veterans' reactions to an offer of psychotherapy. Unpublished doctoral dissertation, Ohio State Univer., 1953.
2. HANFMANN, E. William Stern on "Projective Techniques." *J. Pers.*, 1952, **21**, 1–21.
3. KORNER, A. F. Theoretical considerations concerning the scope and limitations of projective techniques. *J. abnorm. soc. Psychol.*, 1950, **45**, 619–627.
4. LINDZEY, G. Thematic Apperception Test: Interpretative assumptions and related empirical evidence. *Psychol. Bull.*, 1952, **49**, 1–25.
5. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
6. ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.
7. SOSKIN, W. F. Bias in postdiction from projective tests. *J. abnorm. soc. Psychol.*, 1954, **49**, 69–74.

Received February 11, 1957.

# THE REPRESSION HYPOTHESIS STUDIED IN A SITUATION OF HYPNOTICALLY INDUCED CONFLICT

RUTH A. BOBBITT[1]
*State University of Iowa*

In recent years, a number of experimental psychologists have tried to rewrite psychoanalytic theory in operational terms. The concept of repression has stimulated a variety of studies (13, 16, 21, 22, 23, 24, 25) that have succeeded primarily in pointing up methodological problems but have contributed little to our knowledge of repression as either a phenomenon or a process. This paper reports a study made ten years ago (3) that still seems pertinent today.

Repression, as a method of ego defense, protects the individual from painful emotional states by keeping ideas that represent certain impulses out of consciousness (7, p. 86). Following Freud's formulation, we may identify the necessary conditions for the development of repression as these: (*a*) An instinctual impulse is aroused; (*b*) gratification of this impulse is in some way unacceptable to the ego of the individual; (*c*) pain (anxiety) has been experienced as the ego demands have been frustrated in gratifying the instinctual tendency (7, p. 84; 9, p. 119); and (*d*) the need to avoid this anxiety is stronger than the instinctual drive (7, p. 85).

As a process, repression as conceived by Freud has the following dynamic characteristics:

1. The repressed is only separated from consciousness (awareness) and is presumably still active in the unconscious. In fact, the extent and degree of repression varies as inhibitory and instinctual propensities fluctuate (7, pp. 89–90).

2. Generalization of cues and responses often occurs. Cues originally related to the drive may be repressed, but other representations may be admitted to consciousness and acted upon with the accumulated energy of the repressed (7, pp. 86–87). Or, the related ideas may become so associated with the repressed that they must also undergo repression, if the individual is to avoid anxiety (Freud's "repression proper," (7, p. 86). If repressed impulses are satisfied, it must be through substitute activities. Unless such satisfactions are complete, the need persists, to be expressed through an increasing number of substitute ideas and with "the damming-up consequent on lack of real satisfaction" (7, p. 87).

Since "the motive and purpose of repression is simply the avoidance of 'pain'" (7, p. 92), it follows that repression may be considered successful only to the extent that it does avoid anxiety or pain. It has been noted by Freud (10, pp. 21–22) and by Merrill (16) and Zeller (23) that unsuccessful, i.e., partial repression, or repression in which at least the affective component has become attached to a substitute idea is most amenable to analysis or experiment. These emotional components become apparent when one considers the dynamics of psychoanalytic therapy in which the aim is to remove repressions. Freud says, ". . . the patient can go on spinning a whole chain of . . . associations, till he is brought up in the midst of them against some thought-formation, the relation of which to what is repressed acts so intensely that he is compelled to repeat his attempt at repression" (7, p. 88). Finally, summarizing the formula for therapy, ". . . all repressions are to be undone; the mental condition is then the same as if all amnesias are removed. When all gaps in memory have been filled in, all the enigmatic products of mental life elucidated, the continuance and even the renewal of the morbid condition is impossible" (8, p. 269).

In this theoretical context, this investigation is specifically concerned with anxiety behavior as the response variable to be measured as a function of degree (completeness) of repression. In terms of the general hypothesis, and apart from therapeutic intervention, the following relationships are predicted: In the absence of either conflict or repression, there should be no anxiety behavior. In conflict with full re-

[1] The investigation was conducted under the direction of Professor Charles R. Strother; technical assistance was given by Drs. Wilbur R. Miller and Paul Huston of the Psychopathic Hospital, and by Professors Robert R. Sears and E. E. Lindquist of the State University of Iowa. The author now resides at 3900–50 N.E., Seattle, Washington.

pression, there should be no anxiety behavior except as the repressive process fluctuates. With incomplete repression, there should be an increase in the scope and intensity of anxiety disturbance. As repression ceases to operate disturbance should reach a maximum and show considerable generalization. Finally, when both repression and conflict are removed there should no longer be anxiety behavior related to the conflict.

Most of the investigations of repression have utilized some measures of memory as the response variable. Zeller (23, pp. 40–45) recently provided an excellent critical review of these, pointing out the fallacies and inadequacies of each. The kind and amount of feeling tone, emotional disturbance, or ego threat in the experimental situation are seen as important factors. He also stresses that any critical study must demonstrate "restoration to consciousness of the repressed material" upon removal of the repression factor (23, p. 46).

Zeller (23, p. 44) and other writers (16, 22) have noted the Huston, Shakow, and Erickson experiment (13) as one of the better demonstrations of repression, although the use of hypnosis makes interpretation somewhat difficult. The criticism in terms of dependence on hypnosis loses some weight when we consider that results in conformity with the repression hypothesis occur as predicted, and that these results could not have been predicted on the basis of hypnosis alone. The hypnotic approach surely warrants further investigation, if only for the fact that other approaches thus far devised involve variables perhaps even more difficult of appraisal, i.e., learning, motivation, reminiscence, retroactive inhibition, etc. (24, p. 422).

A further consideration in experimental design is the response to be measured as a function of repression. In the case of most of the experiments measuring memory responses, "anxiety" or some form of personal involvement (frustration, incompleteness, unpleasantness) is inferred. In the study of Huston, Shakow, and Erickson (13), however, the responses measured are presumed on the basis of tests of validity to be measures of emotional disturbance or anxiety. The condition of repression is of course also independently validated.

## METHOD

This experiment employs hypnotically induced conflicts as described by Erickson (6) tested according to Luria's association test procedures (13, 15). Experimentally induced conflicts offer opportunity for necessary control. Except when hypnosis has been employed, however, studies of conflict in human subjects (*S*s) have been too circumscribed and moderate for our purposes. The hypnotic method has been found to have some validity, i.e., the experimenter can define the nature of the conflict, assure the *S*'s susceptibility to and acceptance of it, suggest in general the nature of his reaction to it, create the effects of repression and, finally, remove the hypnotic effects without undesirable residual reactions.

Luria's method for testing emotional disturbance calls for conditioning a voluntary motor pressure of the preferred hand to occur simultaneously with a verbal response. The theory is that an integrated response of low habit strength under voluntary control is notably susceptible to disorganizing effects of emotional disturbance. His report of the general validity of the method has been verified in a group of studies (1, 4, 5, 11, 12, 13, 18, 19, 20).

The design of this experiment is as follows: Under hypnosis, an incident created to arouse a severe state of conflict is related to six selected *S*s. In the waking state, responses to a standard set of 100 association test words are recorded under five experimental conditions: (*a*) before the conflict is induced; (*b*) after induction of the conflict and with posthypnotic amnesia for it; (*c*) with partial posthypnotic amnesia; (*d*) with full awareness of the conflict as an immediate actuality; and (*e*) after hypnotic removal of the conflict. The association test is composed of "critical" words relating directly to the conflict story; "postcritical" words immediately following critical words, and "neutral" words to which the *S*'s responses were in no way disturbed on a preliminary testing. The three experimental variables are thus subjects (assumed to contain a random effect), types of verbal stimuli and conditions of conflict awareness. The simple factorial design permits evaluation of the three variables in all combinations by means of analyses of variance.

### *Subjects*

*Ss* were selected from a group of male junior and senior medical students who had no previous experience with hypnosis, but were willing to serve in a reaction time experiment involving hypnosis. After training, each *S* was rated hypnotizable and capable of complete posthypnotic amnesia.[2] No individuals were accepted whose personalities were rated "extreme" on the basis of a brief Rorschach evaluation, a psychiatric interview and behavior during hypnotic training sessions.

### *Procedure*

The same conflict story and instructions governing degree of awareness were related under hypnosis to each *S*. In outline, the story was this:

In medical school, the *S* has had an experience so disturbing that he had to forget it. As it is recalled, he will re-experience it in detail. He was attending a party and volunteered to borrow a car and go for more beer. On his way back, he drove rapidly around a dark corner and felt the car hit something. He feared an accident but did not want to stop since he was in a borrowed car and had been drinking. He rationalized that it had probably been only a bump in the road and drove on. Next morning his fears were confirmed in a newspaper story of a child, near death and hopelessly crippled as the result of a hit and run accident. The details of the account made it clear to him that he was the driver for whom the police were searching. He experienced great panic and conflict about what to do.

Each condition of awareness was induced under hypnosis after review of the story. For the postconflict condition, the hypnotist proved to *S* under hypnosis that the incident could not have occurred and talked freely with him to assure removal of the conflict. After completion of the experiment, a final hypnotic session assured removal of the conflict; it was suggested also that the *S* would recall his thoughts and feelings, wanting to talk freely about them when he awakened.

The association test was administered before hypnosis (Condition I), in the waking state after each hypnotic session (Conditions II, III, and IV), and after the conflict was removed (Condition V).

Independent ratings of each *S*'s behavior and introspective accounts were made by two judges, the hypnotist and the *E*. Unanimous agreement that *S* had accepted the conflict and with respect to levels of awareness was required.

The 100-word association test, composed of 60 neutral, 20 critical, and 20 postcritical words, was individually constructed for each *S*. Neutral and postcritical words were selected on the basis of *S*'s responses to a preliminary test of 210 words. Words were excluded when any measure of response deviated markedly from his mode. Eighty words were then drawn at random from the remaining list. The critical words were either key words in the conflict story (*collision*, *party*, *newspaper*, *borrow*, etc.) or suggested feelings and ideas closely related to it (*criminal*, *guilty*, *conflict*, etc.). These words were randomized and distributed at varying intervals among the neutral words. The same order of words was maintained throughout the testing.[3]

The apparatus employed (2, 3) furnished several response measures: verbal reaction time (VRT), time between beginning of verbal response and beginning of voluntary motor pressure (VR-VMP), number of VMP made during a single stimulus-response sequence (# VMP), duration of the first VMP in a stimulus-response sequence (DVMP), number of involuntary motor pressures occurring in a stimulus-response sequence (# IMP), duration of IMP (DIMP), breathing rate (BR), and number of conflict-associated verbal responses obtained during a single testing period (# CAVR). These measures were reliable, as indicated by agreement between two scorers. Reliability coefficients range from .84 to .997 on measured items and percentages of agreement range from .916 to 1.00 on counted items.

## RESULTS

The validity of the experimental situation, apart from that of the response measures, is assumed on the basis of unanimous judges' ratings and on the consistency among observations and introspective reports. Furthermore, there appears to be descriptive similarity between repression, as postulated here, and the posthypnotic conditions of awareness.

Validity of response measures could be tested more directly. By means of analyses of variance, the three experimental variables (subjects, stimuli, and conditions) were tested in interaction with one another (14). Since none of these tests was significant, the specific differences relating to validity and to the experimental hypotheses were tested by the triple interaction variance as the error term.

Table 1 summarizes results of the validity tests.[4] It is seen that (*a*) all response measures, except BR, tend to differentiate responses in the conditions of conflict from those in conditions of nonconflict (see Lines 1*a*, 1*b*, and 1*c*); (*b*) all measures but DVMP tend to differentiate between responses to critical and to

[2] Hypnosis was conducted by three psychiatrists, Drs. Paul Huston, A. W. Freidinger, and F. S. Bobbitt.

[3] Maintaining the same sequence of words throughout the experiment makes likely the operation of practice effects in opposition to the experimental hypotheses. Changing sequence, however, might introduce factors less subject to appraisal.

[4] Tests significant at or below the .05 level of confidence are considered "significant"; those between the .05 and .20 levels are considered "suggestive." This procedure is used to point up consistent trends. The small number of *Ss* makes necessary large differences, considerable precision and intersubject homogeneity for a high level of statistical significance; requirements difficult of fulfillment in an experiment of this kind.

TABLE 1

RESULTS OF VALIDITY TESTS OF ALL RESPONSE MEASURES SHOWING FOR EACH THE LEVEL OF CONFIDENCE WITH WHICH THE HYPOTHESIS OF NO DIFFERENCE MAY BE REJECTED

| Difference tested | Response Measure[a] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | VRT | VR-VMP | DVMP | ∫VMP | ΔR | ∫IMP | DIMP | CAVR |
| 1. Between means for conflict and nonconflict conditions | | | | | | | | |
| *a*. Critical words, conflict minus critical words, nonconflict | 1% | 2–5% | 5–10% | 10–20% | —[b] | 5–10% | 5–10% | —[b] |
| *b*. All words, conflict minus all words, nonconflict | | | | | 20% | | | |
| *c*. All words, conflict minus all words, postconflict | | | | | | | | 10–20% |
| 2. Between mean response to critical and to neutral stimulus words | | | | | | | | |
| *a*. Critical words, conflict minus neutral words, conflict | 1% | 5–10% | 20% | 10–20% | —[b] | 10–20% | 10–20% | —[b] |
| *b*. Critical words, nonconflict minus neutral words, nonconflict | 20% | 20% | 20% | 20% | —[b] | 20% | 20% | —[b] |

[a] Response measure:
VRT—Verbal reaction time
VR-VMP—Time between verbal response and voluntary motor pressure
DVMP—Duration voluntary pressure
VMP—Voluntary motor pressure
BR—Breathing rate
IMP—Involuntary motor pressure
DIMP—Duration involuntary pressure
CAVR—Conflict associated verbal response

[b] Response measure could not be tested for this difference.

neutral words under the conflict conditions (Line 2*a*). Conversely, no difference between responses to these words is shown in the nonconflict conditions (Line 2*b*). It is interesting to note, in terms of Luria's findings, that VRT and VR-VMP can be considered valid with greater confidence than any of the other measures.

It is now possible to evaluate the findings in relation to the experimental hypotheses. When hypothetical results (Fig. 1) are compared with obtained results for the group (Fig. 2), general conformity is apparent. However, maximum disturbance tends to occur more often in Condition III (partial awareness) than in Condition IV (full awareness), and there is little generalization of disturbance in responses to postcritical and neutral stimuli.

The specific experimental predictions have been subjected to statistical test; reported in Table 2, however, are only those tests warranted by the validity tests (Table 1) and study of Fig. 2. Breathing rate and conflict-associated verbal responses are omitted as are comparisons involving responses to postcritical and neutral words.

1. There was little disturbance under the condition of conflict without awareness (Table 2, Line 1). Only VRT and DIMP tend to differentiate between these conditions. That there is no particular disturbance in relation to other types of stimuli is evident in Fig. 2.

2. There is increased disturbance in response to critical words in the condition of partial awareness over the condition of conflict without awareness (Table 2, Line 2) and over the condition of preconflict (Table 2, Line 3). There is also the predicted involvement of more response measures (Table 2, Line 3 in relation to Line 1), but obviously not a significant generalization of disturbance to postcritical and neutral stimuli (Fig. 2).

3. The predicted further increase (to a maximum) in disturbance under the condition of conflict with full awareness (Table 2, Lines 4 and 5) did not occur. Fig. 2 shows that, for most measures, disturbance in this condition (IV) is somewhere between the other two conflict conditions. For all measures though, responses to critical words are significantly different in the condition of full awareness from those in the preconflict condition (Table 2, Line 6).

4. All response measures lose their disturbed character when conflict is removed but awareness maintained (Table 2, Line 7). No response

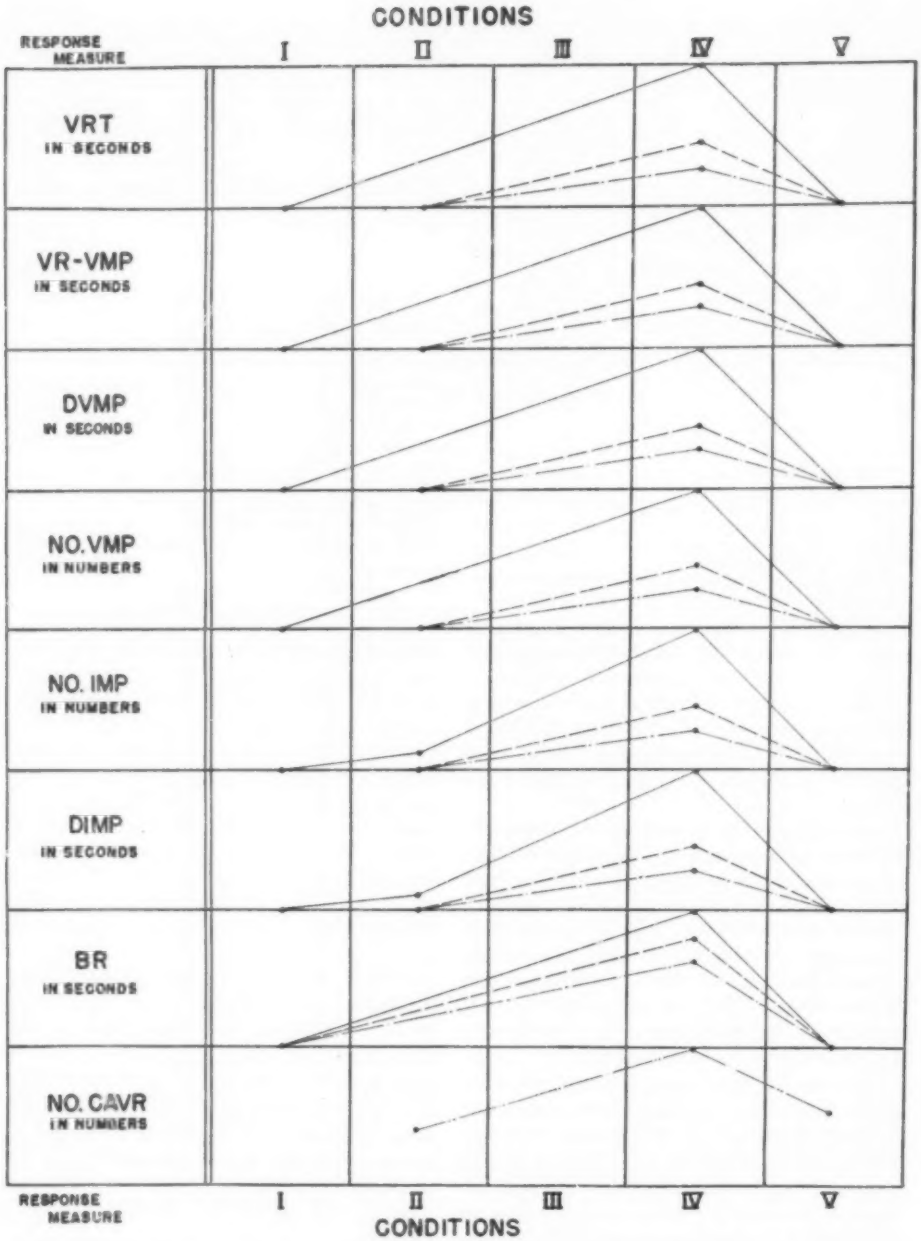


FIG. 1. HYPOTHETICAL CURVES OF RESPONSE MEASURES FOR THE TOTAL GROUP BY EXPERIMENTAL CONDITIONS AND TYPES OF STIMULI

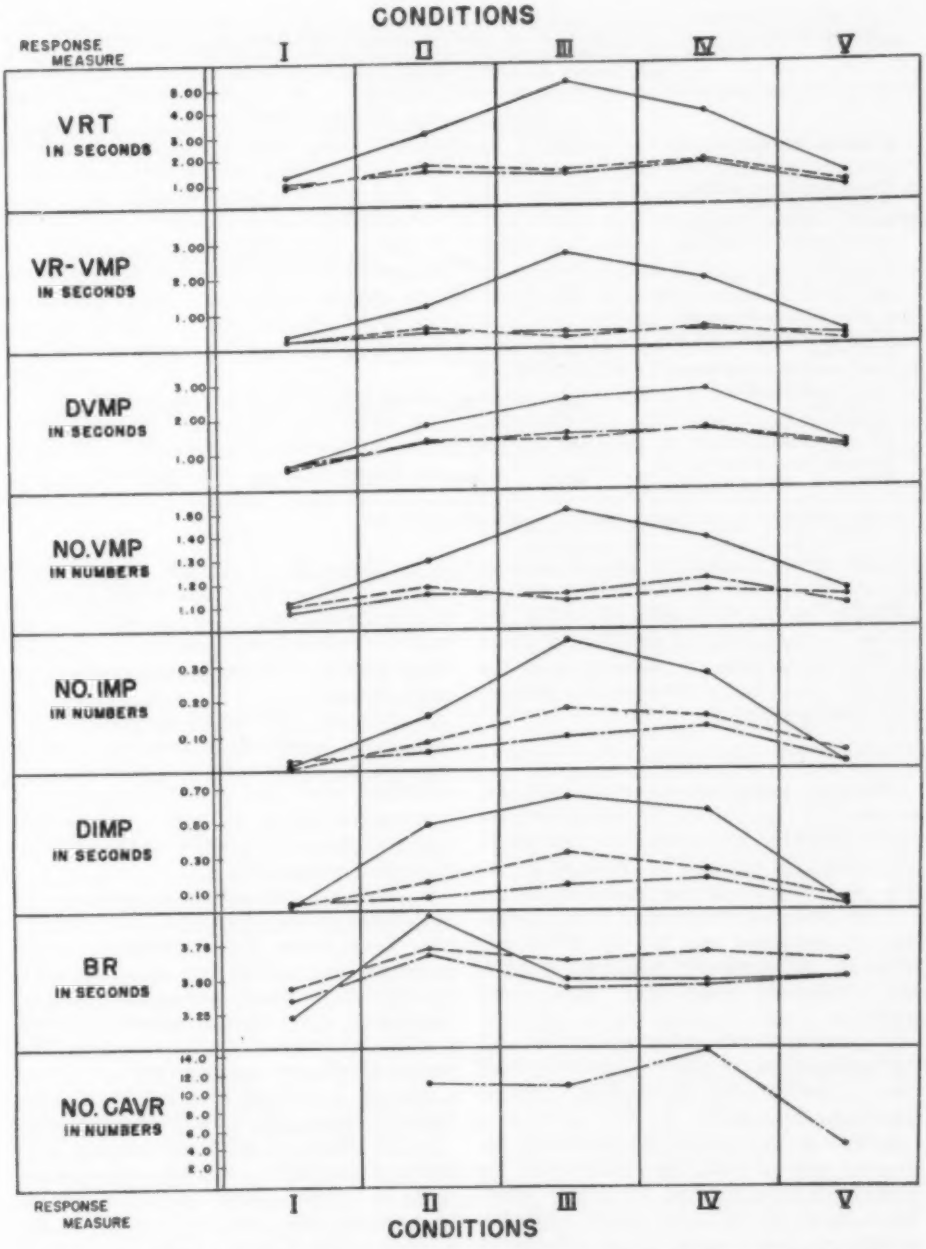


FIG. 2. MEANS OF RESPONSE MEASURES FOR THE TOTAL GROUP BY EXPERIMENTAL CONDITIONS AND TYPES OF STIMULI

TABLE 2

TESTS OF THE "DEGREE OF AWARENESS" HYPOTHESIS SHOWING THE LEVEL OF CONFIDENCE WITH WHICH THE HYPOTHESIS OF NO DIFFERENCE MAY BE REJECTED
(For each response measure, for critical stimulus words only)

| Difference Between Means | Response Measure | | | | | |
|---|---|---|---|---|---|---|
| | VRT | VR-VMP | DVMP | VMP | IMP | DIMP |
| 1. Preconflict minus conflict without awareness | 5–10% | 20% | 20% | 20% | 20% | 10–20% |
| 2. Conflict without awareness minus conflict with partial awareness | 1–2% | 5–10% | 20% | 10–20% | 10–20% | 20% |
| 3. Preconflict minus conflict with partial awareness | 1% | 1% | 2–5% | 1–2% | 1% | 2–5% |
| 4. Conflict with partial awareness minus conflict with full awareness | 10–20%[a] | 20% | 20% | 20% | 20% | 20% |
| 5. Conflict without awareness minus conflict with full awareness | 20% | 20% | 10–20% | 20% | 20% | 20% |
| 6. Preconflict minus conflict with full awareness | 1% | 2–5% | 1–2% | 10–20% | 2–5% | 5–10% |
| 7. Preconflict minus postconflict | 20% | 20% | 20% | 20% | 20% | 20% |

[a] This tendency is for verbal reaction time to be significantly less disturbed.

measure differs significantly between preconflict and postconflict conditions. With no data available on conflict-associated verbal responses in the preconflict condition, it is not possible to evaluate the significance of the average of four such responses in the postconflict condition (Fig. 2).

## DISCUSSION

Two discrepancies, one major and one minor, between the experimental hypotheses and obtained data have been noted. First, maximum disturbance did not occur in Condition IV. In this connection, it must be considered that:

1. All obtained differences between conditions of awareness may be due to chance. However, the consistent statistical, graphic, and introspective results make this a remote possibility.

2. Contrary to the judges' evaluations and the introspective accounts, there may have been no real difference in awareness between Conditions III and IV.

3. With development of full awareness, the situation may have had less conflict-value. No *S* expressed doubt that the accident occurred. Three *S*s stated that they feared inevitable punishment and show a general downward trend in disturbance. Two *S*s had decided against conscience and in favor of self-protection. One of these showed greatest disturbance in Condition IV, while the other showed somewhat inconsistent increases and decreases in disturbance. One *S*, whose behavior was much the same in Conditions III and IV, could see no solution to the conflict but was trying to avoid detection.

If there was resolution of conflict in Condition IV, the hypothesis as formulated may be true, but most of the test may have been administered after resolution began and may measure the process of conflict and anxiety reduction instead of reaction to the development of full awareness.

4. The data may reflect an error in the formulation of the hypothesis concerning degree of awareness. The development of full awareness of a conflict may be accompanied by diminution in disturbance so that maximum disturbance occurs when awareness is at least partial but somewhat less than complete. The peak of disturbance may come with developing awareness, but almost immediately other defenses operate to reduce disturbance (anxiety).

5. The effects of full awareness may have been counteracted by some habituation factor. Tests of practice effects, i.e., differences between Conditions I and V, do not show significant results (Table 2, Line 7). Tests of abreaction or desensitization cannot be made separate from those of the experimental hypotheses.

The second discrepancy is that the expected generalization of disturbance to postcritical and neutral stimulus words did not materialize to any significant extent. The qualitative findings of Luria and of Huston, Shakow, and Erickson (13, 15) suggest such generalization, while the statistical findings of Gardner (11) do not. Generalization may depend on several factors, two of which should be mentioned here:

1. The time interval between a verbal response and the succeeding verbal stimulus (in this study, approximately three seconds were allowed to elapse to permit dissipation of disturbance in the appropriate stimulus-response sequence).

2. The remoteness of postcritical and neutral words from critical subject matter (in this experiment, rather unusual care was taken, in selecting neutral words, to avoid critical implications).

## SUMMARY

Based on the Freudian concept of repression, an experiment was designed to test the following specific hypotheses with respect to anxiety behavior:

1. Repression serves to protect the individual from anxiety related to the repressed impulse.

2. The less successful the repression, the greater is the anxiety experienced in relation to conflict between the repressed and the ego demands.

3. Complete failure of repression results in maximum anxiety, a condition which persists until a new repression is instituted or another defense becomes operative.

4. Complete relief from both repression and conflict results in cessation of anxiety.

Through hypnotic induction and control of conflict and degree of awareness, five experimental conditions were produced: (*a*) preconflict control; (*b*) conflict with posthypnotic amnesia for the situation; (*c*) conflict with partial posthypnotic amnesia; (*d*) conflict with full awareness; and (*e*) postconflict control. Eight measures of response to a word-association test administered by the Luria method were obtained for six *S*s. Each *S* was tested in the waking state under the five experimental conditions.

For the *S*s as a group it was demonstrated that (*a*) there was a conflict; (*b*) there was a progressive increase in degree of awareness; and (*c*) most of the measures of behavior reflected emotional disturbance with some validity.

Results were as predicted except in two respects:

1. There was not the anticipated maximum disturbance when *S*s were made fully aware of the conflict. It is suggested that (*a*) with failure of repression certain defenses quickly operate to diminish anxiety, or (*b*) the hypothesis requires reformulation at this point.

2. Predicted generalization of disturbance to neutral stimulus material did not occur, although there was generalization to an increasing number of response measures. It appears that the testing procedure allowed most disturbances to dissipate in the stimulus-response sequence in which they began.

## REFERENCES

1. BARNACLE, C. H., EBAUGH, P. G., & LEMERE, F. Association motor investigation of the psychoneuroses. *Amer. J. Psychiat.*, 1935, **91**, 925–937.
2. BOBBITT, RUTH A. A new apparatus for the Luria experiment. *J. exp. Psychol.*, 1940, **27**, 578–582.
3. BOBBITT, RUTH A. An experimental investigation of certain aspects of the repression hypothesis. Unpublished doctoral dissertation, State Univer. of Iowa Library, 1947.
4. BURTT, H. E. Motor concomitants of the association reaction. *J. exp. Psychol.*, 1936, **19**, 51–63.
5. CROSLAND, H. R. The psychological methods of word-association and reaction-time as tests of deception. *Pub. Univer. Oregon, Psychol. Series I*, 1929, **1**, 104.
6. ERICKSON, M. H. The method employed to formulate a complex story for the induction of an experimental neurosis in a hypnotic subject. *J. gen. Psychol.*, 1944, **31**, 67–84.
7. FREUD, S. Repression. *Collected Papers*, 1915, **4**, 84–97.
8. FREUD, S. Early papers and history of psychoanalytical movement. *Collected Papers*, 1924, **1**, 379.
9. FREUD, S. *New introductory lectures in psychoanalysis*. New York: W. W. Norton, 1933.
10. FREUD, S. *The problem of anxiety*. New York: Psychoanalytic Quarterly Press & W. W. Norton, 1936.
11. GARDNER, J. W. An experimental study of the Luria technique for detecting mental conflicts. *J. exp. Psychol.*, 1937, **20**, 495–506.
12. HOUTCHENS, H. M. A study of mental conflict in delinquent and nondelinquent boys. Master's thesis, Univer. of Iowa, 1935. (Abridged in *J. Juv. Res.*, 1935, **19**, 180–192.)
13. HUSTON, P. E., SHAKOW, D., & ERICKSON, M. H. A study of hypnotically induced complexes by means of the Luria technique. *J. gen. Psychol.*, 1934, **11**, 65–97.

14. LINDQUIST, M. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.
15. LURIA, A. *The nature of human conflicts.* (Trans. by W. Horsley Gantt.) New York: Liveright, 1932.
16. MERRILL, R. An experimental study of repression. Unpublished doctoral dissertation, Univer. of Washington Library, 1953.
17. MILLER, N. E., & DOLLARD, J. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
18. MORGAN, M. I., & OJEMANN, R. H. A study of the Luria method. *J. appl. Psychol.*, 1942, **26,** 168–179.
19. OLSON, D. M., & JONES, V. An objective measure of emotionally toned attitudes. *J. genet. Psychol.*, 1931, **39,** 174–196.
20. REYMERT, M. L., & SPEER, G. S. Does the Luria technique measure emotion or merely bodily tension? *Character & Pers.*, 1939, **7,** 192–200.
21. ROSENZWEIG, S. Investigation of repression as an instance of experimental idiodynamics. *Psychol. Rev.*, 1952, **59,** 339–345.
22. SEARS, R. R. Functional abnormalities of memory with special reference to amnesia. *Psychol. Bull.*, 1936, **33,** 229–274.
23. ZELLER, A. F. An experimental analogue of repression. I. Historical summary. *Psychol. Bull.*, 1950, **47,** 39–51
24. ZELLER, A. F. An experimental analogue of repression. II. The effect of individual failure and success on memory measured by relearning. *J. exp. Psychol.*, 1950, **40,** 411–422.
25. ZELLER, A. F. An experimental analogue of repression. III. Effect of induced failure and success on memory measured by recall. *J. exp. Psychol.*, 1951, **42,** 32–38.

Received April 22, 1957.

# THE EFFECT OF SIBLING COMPETITION AND ALLIANCE ON LEVEL OF ASPIRATION, EXPECTATION, AND PERFORMANCE[1]

EVA DREIKURS FERGUSON

*Western Psychiatric Institute, University of Pittsburgh*

THE present study seeks to test certain hypotheses that have arisen from the Adlerian theory of sibling relationships. According to this theory, personality patterns are a reflection of goals that the individual establishes within the family constellation of the early years of childhood. A goal basic to all human behavior is the wish to belong, to be accepted (1, 2, 4), so that the child's behavior patterns are expressions of trying to find a place, or status, within the family group. The presence of siblings in the family can be viewed by the child as an obstacle to attaining status, in which case the relationship he establishes with his siblings is a competitive one. If, however, the sibling is viewed as an ally in the process of gaining status in the family, the interaction between these two children takes on a pattern different from that of the competitive relationship. Since early childhood strivings for status find expression in individual personality development, the personality patterns of the siblings reveal the type of relationship established between them: allies become alike in personality, competitors become different. Dreikurs, in expanding the original conceptualization of Adler, has described as follows the dynamics of the sibling interaction and their effects on personality development:

> As each member of the family strives for his own place within the group, the competing opponents watch each other carefully to see the ways and means by which the opponent succeeds or fails. Where one succeeds, the other gives up; where one shows weakness or deficiencies, the other steps in. In this way competition between two members of the family is always expressed through differences in character, temperament, interests, and abilities. Conversely, the similarity of characteristics always indicates alliances. Sometimes, the two strongest competitors show no sign of open rivalry but, rather, present a close-knit pair; nevertheless, their competitive striving is expressed in personality differences (p. 40, 4).

Competition, as used here, thus describes a specific behavior pattern: giving up where the other person succeeds and striving to outdo only where the other fails, with the result that competitors display contrasting personality patterns. Dreikurs differentiates rivalry from competition (4, 5), rivalry being a struggle for superiority in specific situations or for limited goals, while competition is a pervasive and long-range struggle over the attainment of status within the family group. Moreover, rivalry need not display the "seesaw" phenomenon postulated above for competition. According to this distinction, rivalry may occur between any two individuals, including allies.

This conceptualization of competition and its effect on personality development has found validation in clinical practice, as for example in therapy with "problem children": once the problem child improves with therapy, his previously "good" competitor usually displays problem behavior that is entirely novel for him. Rigorous experimental tests of these formulations have not yet been conducted, however. The present study therefore attempted to clarify to what extent siblings who are Allies and those who are Competitors, as defined by sameness or oppositeness of traits, do behave in the manner described by Dreikurs. For this purpose, sibling pairs were given three experimental tasks, and two types of response measures were utilized: (*a*) performance and (*b*) statements of aspiration. Two hypotheses were derived from Dreikurs' position:

*Hypothesis 1.* In a competitive relationship, a child's performance, level of aspiration, and level of expectation ***rise*** when he is informed that his sibling has done more poorly than himself and ***drop*** when he is told that his sibling has done better than himself.

*Hypothesis 2.* In an alliance relationship, a child's performance, level of aspiration, and level of expectation ***drop*** when he is informed that his sibling has done more poorly than himself and ***rise*** when he is told his sibling has done better than himself.

[1] This study is a condensation of the doctoral dissertation (6) conducted in partial fulfillment for the Ph.D. degree at the Department of Psychology, Northwestern University. Grateful acknowledgments are extended to Drs. Benton J. Underwood, dissertation supervisor, and Rudolf Dreikurs, who served as interviewer and gave invaluable suggestions.

On the basis of data to be presented, it was found that these hypotheses required substantial modification. Differences did emerge between Competitors and Allies, but not as hypothesized nor did they occur on all tasks and response measures. Upon *post hoc* consideration of the task dimensions in the present study, it seemed necessary to take into account two additional variables, ego-involvement and status-effect, in framing a more adequate test of the Adler-Dreikurs formulation. These variables are here so defined as to link on a common dimension ego-involvement and success-failure evocation, variables which in the literature are largely treated as discrete phenomena and which to the present writer appear conceptually related. *Ego-involvement* is here defined as occurring when *S* perceives the task or situation as important to the maintenance or raising of his status, either because of the nature of the task or because of instructions given to *S*.[2] *Status-effect* is defined as occurring when *S* experiences a *shift* in status. Thus, under conditions of failure *S* experiences a decrease of status, and under conditions of success *S* experiences an increase of status. Presumably, ego-involvement (calling status into play) heightens status-effect, although operationally ego-involvement and status-effect are differentiable. Both for the purpose of the present study, as well as for a clarification of these problems in the literature, this kind of distinction may be found useful.

Although ego-involvement and status-effect were not varied systematically in the present study, *post hoc* analysis of the experimental procedure and of the obtained data suggested a revised formulation that gives prominence to these two variables. This formulation is presented at this point to facilitate the exposition of results along lines believed by the investigator to be promising for further research. A basic hypothesis in the revised conceptualization is that differences between Allies and Competitors appear in a sibling comparison situation when ego-involvement and status-effect are present to a high or moderate degree but not when they are minimally present. When these variables are present to a *high* degree, as they are in much of the real-life experience of the child, Allies and Competitors may be expected to behave as predicted by the original hypotheses, i.e., Allies strive to do like the sib, Competitors give up when the sib does better and only strive to outdo the sib when the sib does worse. When ego-involvement and status-effect are *moderate*, however, Competitors behave in a manner analogous to rivalry while Allies are relatively unaffected by the sib comparisons, being more attentive to their own than to their sib's performance. This revised hypothesis leads to several ramifications. For one thing, it suggests that the alliance relationship is a weaker one than that of competition, stronger ego-involvement and status-effect being necessary for the Allies than for the Competitors to show a change in a situation of sib comparison. It further suggests that the type of behavior shown by Competitors is contingent upon whether the sib comparisons are presented in a context of moderate or of high ego-involvement and status-effect. Finally, the revised formulation does not predict difference between Allies and Competitors when presented with sib comparisons under conditions of *low* ego-involvement and status-effect, even though both groups may use knowledge of the sib's performance as cues for their own performance when they lack prior experience with the task. Goal-setting statements, possibly tapping motivational variables more directly than do performance measures, should show response changes more clearly than performance measures in the sib comparison situation. It may therefore be hypothesized that whereas performance and goal-setting statements differentiate the behavior of Allies and Competitors under conditions of high ego-involvement and status-effect, only goal-setting statements do so under moderate ego-involvement and status-effect.

## METHOD

### *Subjects*

Two duplicate experiments were performed. Center *S*s came from families where at least one sib was brought, within the past 18 months, to a child guidance center[3] as a behavior problem; Normal *S*s were selected

[2] This definition of ego-involvement covers more aspects than are usually described in the literature (e.g., 7, 8), and as such appears to have more face validity. It includes raising of status, rather than merely threat to the maintenance of status, and it defines status both in terms of self-esteem and social-esteem, for want of a satisfyingly clear delineation between the two types of esteem.

[3] The clinics are part of the Community Child Guidance Centers of Chicago, a nonprofit organization under the sponsorship of the Individual Psychology Association of Chicago.

from churches in the community in which the university is located. Each experimental group was composed of 20 sib pairs, 10 of Allies and 10 of Competitors, giving a total of 80 *S*s or 40 sib pairs. All *S*s came from families that had three or more children, *S*s being between 4½ and 10 years of age. Allies and Competitors were matched by pairs for age and where possible for sex, approximately half the *S*s in each group being under eight years of age and half being eight years or older. The *S*s were selected on the basis of information obtained in a brief interview with the mother of each pair of sibs. The interviewer was a prominent Adlerian psychiatrist who was instructed to perform the interviews in the manner usual in his psychiatric practice. His diagnoses based on the interview data were found, on analysis subsequent to the completion of the study, to follow a quantitative system, although he was not aware of this and followed what he considered was an intuitive approach in arriving at the diagnoses.

A pre-experimental test of the reliability of the interview method was performed. In this, the interviewer talked to twenty sets of parents without knowledge of which father and mother belonged together. To assess degree of agreement between the diagnoses derived from interview with the father and from interview with the mother, a 3 × 3 chi-square contingency table was prepared for the three diagnostic categories of Allies, Mixed, and Competitors. This yielded a contingency coefficient of .640, significant beyond the .01 level. (This coefficient is to be compared to the maximum possible for a 3 × 3 table, of .816.) In the interviews, the examiner first asked the parent which of the children are most different and which most alike, and in what regard. He then asked the parent to describe each child. If more information was needed, the interviewer asked about the following types of traits: who has the most (least) temper; who minds best (least); who is the best (worst) student; who is the most (least) outgoing.

For the 80 *S*s used in the experiment, *E* did not know the diagnosis of *S*s until the completion of the study, at which time the interview material was analyzed. This analysis revealed that when sibs showed a ratio of at least three traits opposite to one trait alike, the diagnosis was that of competition, and when the trait ratio of alike to opposite was at least three to one, the diagnosis was that of alliance.[4] The 40 experimental pairs were drawn from a total of 104 diagnosed pairs, of which 10 pairs were not usable because one or both of the sibs were outside the required age range. The diagnostic distribution of the 104 pairs was 54 pairs of Competitors, 30 pairs of Allies, and 20 pairs of Mixed diagnosis. Selection of the final sample of 40 pairs was dictated by the requirements of matching Allies and Competitors on a pair basis for age and where possible, for sex. Table 1 presents relevant characteristics of *S*s.

TABLE 1
CHARACTERISTICS OF SIBLING PAIRS

| Pair | Normal Sample | | Center Sample | |
|---|---|---|---|---|
| | Competitors | Allies | Competitors | Allies |
| Number of Pairs | | | | |
| $F_1 - F_2$[a] | 4 | 3 | 1 | 5 |
| $F_1 - M_2$ | 2 | 3 | 2 | 1 |
| $M_1 - F_2$ | 2 | 2 | 3 | 1 |
| $M_1 - M_2$ | 2 | 2 | 4 | 3 |
| Total | 10 | 10 | 10 | 10 |
| Years[b] | | | | |
| Mean age | 7.23 | 7.20 | 7.60 | 7.68 |
| Mean age difference | 3.40 | 3.75 | 2.60 | 2.95 |
| Mean Ratio | | | | |
| Mean ratio of similar to opposite traits | .67/4.20 | 4.08/.35 | .07/3.47 | 4.13/.57 |

[a] $F_1 - F_2$ is to be read that both older and younger siblings are female; $F_1 - M_2$, that the older sibling is female, the younger, male; etc.

[b] Years are computed in half-year units.

## *Tasks*

Two aspiration tasks and one verbal learning task were administered. Both *S*s of the sib pair came to the laboratory at the same time, each *S* being put in a separate room. Each task was presented alternately to each sib, and *S*s had no contact with one another throughout the experiment. The only information each sib had of the other's performance was that supplied by *E*. While *E* tested one *S*, the sib played with an assistant in another room. Time intervals between tasks were arranged to control for warm-up and transfer effects and to make credible the information given to *S* regarding the sib's performance. To ensure *S*s' cooperation throughout the duration of the experiment, all *S*s were told prior to entering the laboratory that they would play some games with *E*, and for being so nice and coming to play the games *S* would get a small prize afterwards. So as not to interfere with the experimental conditions (i.e., the sib comparisons made to *S*) only a minimal amount of encouragement was given to *S*s throughout the experiment.

The first task presented to each *S* (Aspiration Task A) consisted of 40 pairs of dots. The *E* demonstrated how the dots could be connected, and then told *S* that within one minute most children of his and his sib's age could finish putting lines through the dots for 20 pairs of dots, but that the sib was real fast (slow), however, and could finish 30 (or 10) pairs of dots before time was called. Each *S* was asked how many pairs of dots he would *like* to finish (Level of Aspiration, here designated LA) and how many he *thinks* he can finish (Level of Expectation, here designated LE). Half the sib pairs were given the 30 comparison and half the 10 comparison. After stating his LA and LE, *S* was told that this was just a practice game and that he could finish it on completion of all the games. *S* was not allowed to perform on this task so that performance and aspiration on the neutral trials of subsequent tests would not be interfered with. The purpose of Aspiration Task A was to determine the LA and LE under the conditions of *S*'s having no previous knowledge of his own skill.

Subsequent to Aspiration Task A, *S* was shown the first list of Verbal Learning Task B. This task consisted of three lists of pictures, the first six pictures in a list being presented to those *S*s under eight years of age, all eight pictures in a list being presented to *S*s eight

[4] For a fuller account of the interview method, the testing of its reliability, and of the system of weighting traits to arrive at the trait ratios, the reader is referred to the original doctoral dissertation (6).

years and older. The pictures were India-ink tracings of common objects, chosen from the Stanford-Binet Pictures Vocabulary series (Year II–IV, Form L) and from a series used by Chapman (3). Each list of pictures was presented to *S* for serial learning, the intratrial interval being five sec. and the intertrial interval ten sec. The two response measures obtained for each list on this task were number of trials to learn to one perfect trial (a perfect trial being the correct anticipation of each picture in the list within the five-sec. intratrial interval), and number of errors made during the learning trials. List 1 was presented as a neutral list, while Lists 2 and 3 were preceded by information given to *S* regarding his sib's performance. Half the sib pairs were told before being shown List 2 that the sib had done better than *S* in learning the previous list, and told before List 3 that the sib had done worse. The reverse sequence of sib comparisons was given to the remaining sib pairs. When all three lists were presented under the neutral condition in a pre-experimental test with children aged five to ten years, the means and *SD*s of the three lists were found to be comparable.

The third task, Aspiration Task C, was a pegboard with 11 rows of 60 holes each. For each trial, *S* had to insert as many pegs in a row as he could within 60 sec., using only one hand and picking up one peg at a time. *S* was given two practice trials and nine test trials. A trial consisted of *S*'s stating his LA and LE and then filling in the holes. After three test trials, and preceding Trials 4 through 6, half the sib pairs were told the sib had done better than *S*; preceding Trials 7 through 9 these *S*s were told the sib had done worse than *S*. The reverse sequence of sib comparisons was given to the remaining sib pairs. Standardization of this task under neutral conditions with children aged five to ten years, showed performance on the test to be reliable (the obtained Kuder-Richardson *r* being .972) and to suffer from only minimal practice effects.

The twenty *S*s used in standardizing the latter two tasks were drawn from pediatric services at a large research hospital (Illinois Research Hospital, part of the University of Illinois medical school). All *S*s, prior to their participation in the present study, were judged by one or more of the medical or nursing staff to be of at least average intelligence and free of gross personality maladjustment. They were either nonbedridden inpatient convalescents (in a ward run as a kind of playground, so that personality effects from hospitalization appeared to *E* minimal for these *S*s) or outpatients coming for their routine pediatric checkup. From their hospital records and their behavior in the experimental situation, these *S*s appeared to be adequately comparable to the experimental *S*s.

### *Procedure*

In the first task, the "sib comparisons" were statements telling *S* how the sib did in comparison to a *norm*, while in the other two tasks the comparisons stated how the sib did in relation to *S*. Although the wording of the original hypotheses covered specifically only the latter situation, for purposes of study the former condition was included, and in subsequent discussion "sib comparisons" will be used to describe both types of comparisons. In summary, the procedure was as follows:

1. *E* played 2 min. with $S_2$.
2. $S_1$ was given Aspiration Task A (sib completing 10 (30) pairs of dots).
3. $S_2$ was given Aspiration Task A (sib completing 10 (30) pairs of dots).
4. $S_1$ learned List 1 of Verbal Learning Task B (neutral condition).
5. $S_2$ learned List 1 of Verbal Learning Task B (neutral condition).
6. $S_1$ was given Aspiration Task C (three neutral trials, three trials sib doing better [worse], and three trials sib doing worse [better]).
7. $S_2$ was given Aspiration Task C (three neutral trials, three trials sib doing better [worse], and three trials sib doing worse [better]).
8. $S_1$ learned List 2 of Verbal Learning Task B (sib doing worse [better]).
9. $S_2$ learned List 2 of Verbal Learning Task B (sib doing worse [better]).
10. $S_1$ learned List 3 of Verbal Learning Task B (sib doing better [worse]).
11. $S_1$ did Aspiration Task A if he wished, was praised and given his prize.
12. $S_2$ learned List of Verbal Learning Task B (sib doing better [worse]).
13. $S_2$ did Aspiration Task A if he wished, was praised and given his prize.

Comparison conditions were selected for each task by the previous assignment of each sib pair to one of two sequences: (*a*) the 10-Better-Worse sequence and (*b*) the 30-Worse-Better sequence. In the type "*a*" sequence, *S* was told the sib finished 10 dots on Aspiration Task A, the sib did better on Trials 4 through 6 in Aspiration Task C (and worse on Trials 7 through 9) and the sib did worse on List 2 in Verbal Learning Task B (and better on List 3). The type "*b*" sequence is the complete reverse of the type "*a*" sequence. The sequence of comparisons between tasks was thus balanced, i.e., when the comparison was favorable for the sib in the first task (sib finishing 30 dots on Aspiration Task A), the next comparison which *S* received was favorable to *S* (sib doing worse on Trials 4 through 6 on Aspiration Task C), and comparison on the final task was initially favorable again for the sib (sib doing better on List 2 on Verbal Learning Task B). Half each of the Center Competitors, Center Allies, Normal Competitors, and Normal Allies were assigned to the "*a*" condition and the other half to the "*b*" condition. Within each group, the assignment of a given sib pair to the "*a*" or "*b*" condition was random, the only limitation being that no more than half of a given group would fall into one or the other sequence condition.

### *Response Measures*

It will be recalled that in Aspiration Task A no performance measures were obtained and that only one trial was presented to *S*. LA and LE *raw* scores are thus the data used for this task.

For Verbal Learning Task B no aspiration measures were obtained, the two performance measures being number of trials to learn to criterion and number of errors made during the learning trials. The data used for this task were *change* scores from the neutral condition, computed by subtracting the score on the

comparison list from the score on the neutral list. A plus score thus indicates improved performance (fewer errors made, or less trials taken, on the comparison list than on the neutral list), a minus score indicating decreased learning efficiency.

In Aspiration Task C, three response measures were obtained: performance (number of pegs inserted within one min.), LA, and LE. As for Verbal Learning Task B, data for this task are *change* scores, a plus score indicating improved performance or raised aspirations. The change score in Task C was computed by subtracting the mean of the three neutral trials from the mean of the three comparison trials.

## RESULTS

### *Aspiration Task A*

As seen from Table 2, no significant differences were obtained between Competitors and Allies, although both groups had significantly higher LE when told the sib finished 30 pairs of dots than when told the sib finished 10. Without prior experience in a task, thus, both groups used the sib's performance as a reference point by which to set their own expectations. Their aspirations did not follow this pattern, however. The LA scores tended to be uniformly high, except for the Center Competitors who were told the sib finished 10 dots, the low scores of this group primarily accounting for the significant LA interaction found for Center *S*s. The fact that this significant interaction was not found for the Normal *S*s but only for the Center *S*s, can possibly be explained as a function of the counseling experience Center *S*s have been exposed to.

### *Verbal Learning Task B*

As seen from Table 3, no significant effects were found for this task.

### *Aspiration Task C*

No significant effects were found for the performance measure, although both LA and LE measures yielded significant differences between Allies and Competitors. As seen from

TABLE 2
MEAN LA AND LE SCORES ON ASPIRATION TASK A
(*N* per cell = 10)

| Response Measure | Group | Normal Sample | | | | | Center Sample | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sib "Fast" (A) | | Sib "Slow" (B) | | *F* | Sib "Fast" (A) | | Sib "Slow" (B) | | *F* |
| | | Mean | *SD* | Mean | *SD* | | Mean | *SD* | Mean | *SD* | |
| LA | Competitors | 27.9 | 12.60 | 31.0 | 6.63 | | 33.5 | 6.73 | 22.9 | 10.47 | |
| | Allies | 30.0 | 8.94 | 25.5 | 10.59 | | 31.0 | 9.16 | 33.5 | 6.87 | |
| | (Between A & B) | | | | | .04 | | | | | 1.27 |
| | (Interaction) | | | | | 1.31 | | | | | 7.77** |
| LE | Competitors | 24.2 | 9.15 | 19.3 | 5.29 | | 31.3 | 6.80 | 19.8 | 9.17 | |
| | Allies | 24.8 | 7.26 | 17.1 | 5.10 | | 26.6 | 6.86 | 24.3 | 11.72 | |
| | (Between A & B) | | | | | 7.46** | | | | | 5.44* |
| | (Interaction) | | | | | .37 | | | | | 2.42 |

* $p < .05$.
** $p < .01$.

TABLE 3
VERBAL LEARNING TASK B: MEAN CHANGE SCORES IN NUMBER OF TRIALS TO LEARN TO CRITERION AND IN ERRORS
(*N* per cell = 20)

| Response Measure | Group | Normal Sample | | | | Center Sample | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Sib "Better" | | Sib "Worse" | | Sib "Better" | | Sib "Worse" | |
| | | Mean | *SD* | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| Trials[a] | Competitors | −.20 | 2.36 | +.05 | 2.50 | +.80 | 2.29 | +1.25 | 3.00 |
| | Allies | +.65 | 2.48 | +.25 | 2.74 | +.80 | 3.30 | +.25 | 3.71 |
| Errors[a] | Competitors | −2.10 | 6.49 | −1.00 | 6.07 | −.20 | 6.24 | +1.50 | 8.07 |
| | Allies | +1.35 | 6.99 | +.05 | 8.34 | +2.00 | 9.62 | −1.70 | 9.45 |

[a] None of the obtained *F* ratios were significant.

TABLE 4
MEAN CHANGE SCORES ON ASPIRATION TASK C
($N$ per cell = 20)

| Response Measure | Group | Normal Sample | | | | Center Sample | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Sib "Better" | | Sib "Worse" | | Sib "Better" | | Sib "Worse" | |
| | | Mean | *SD* | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| Performance | Competitors | +.60 | 1.51 | +.73 | 1.92 | +1.10 | 1.18 | +1.25 | 1.51 |
| | Allies | +.68 | 1.49 | +.33 | 1.58 | +.87 | 2.66 | +.58 | 2.46 |
| | (Between Competitor-Ally) | $F$ = .18 | | | | $F$ = .96 | | | |
| LE[a] | Competitors | +.60 | 3.85 | +1.05 | 5.16 | +.27 | 3.37 | +1.15 | 3.84 |
| | Allies | +.07 | 5.72 | −3.38 | 8.44 | −.37 | 9.26 | −1.38 | 9.05 |
| | (Between Competitor-Ally) | $\chi^2$ = 2.45 | | | | $\chi^2$ = 5.02* | | | |
| LA | Competitors | +3.10 | 8.75 | +2.48 | 7.63 | +3.02 | 5.88 | +.40 | 7.90 |
| | Allies | −.30 | 8.05 | −2.15 | 9.59 | −3.22 | 8.14 | −2.38 | 5.26 |
| | (Between Competitor-Ally) | $F$ = 4.21* | | | | $F$ = 8.09** | | | |

* $p < .05$.
** $p < .01$.
[a] Both for Normal and Center LE, Bartlett's test for homogeneity of variance gave $\chi^2$ significant beyond the .02 level.

Table 4, the Allies lowered and the Competitors raised their LA and LE under both the Better and Worse comparison conditions. Since heterogeneity of variance was found for the LE measure, a chi-square test suggested by Wilson (9) was used for this measure. This yielded a significant difference in LE between Competitors and Allies for only the Center *S*s. Since the results were substantially in the same direction for both Center and Normal *S*s, however, the conclusion is warranted that both LA and LE differentiated between the responses of Competitors and Allies when *S*s were presented with sib comparisons.

When the data of Table 4 are further analyzed, the overall rise in score from the neutral condition seen in the Competitors' LA and LE data turns out to obscure a more complex relationship. As shown in Table 5, for Competitors a significant interaction is obtained between Better-Worse and First-Second presentation on both the LA and LE measures. This interaction describes the fact that the two conditions, of *S* first being told his sib did better and of *S* secondly being told his sib did worse, produce a significantly greater rise in LA and LE than do the two conditions of *S* first being told the sib did worse and of *S* secondly being told his sib did better.[5] This interaction pattern did not appear for the Allies. The presence of this pattern for Competitors suggests that the behavior of the Competitors may be characterized as largely that of rivalry: when Competitors hear that the sib does better, a set to outdo the sib is elicited and maintained even when subsequently the sib does worse, this set to outdo the sib appreciably not being manifested when *S*s first are told the sib does worse.

The meaning of the Allies' drop in LA and LE scores on being presented with the sib comparisons can probably be best understood by inspection of the data of *S*s used to standardize this task. As will be recalled, these *S*s simply took their nine test trials in the manner of the neutral trials of the experimental *S*s, i.e., no comparisons being presented. These standardization *S*s dropped their LA and LE scores from the first three trials to each set of subsequent three trials. In other words, merely performing the task lowers the LA and LE in *S*s of this age, the goal-setting statements becoming closer to the actual performance as a function of increased experience in the task. As seen in Table 6, Allies are comparable to the standardization *S*s in their drop of LA and LE from the first three trials to the subsequent sets of three trials, the introduction of the sib

[5] Since *S*s were randomly assigned to the sequence conditions and since the pattern is similar for Center and Normal *S*s, a selective *S* variable to account for the significant interaction (i.e., that *S*s in the Better First-Worse Second sequence are somehow different from *S*s in the Worse First-Better Second sequence) does not appear to be tenable.

comparisons after the first three trials producing behavior in the Allies that is substantially no different from the behavior of *S*s who were given no comparisons of any kind. This picture is not found for the Competitors, whose change in LA and LE scores from the first three trials differs significantly from the change scores obtained for the standardization *S*s. From this analysis it can be reasonably concluded that Allies, in contrast to the Competitors, were largely unaffected by the sib comparisons, paying heed primarily to their own performance in a manner analogous to *S*s who never were presented with any comparisons. The Allies' lack of apparent concern over the sib's performance suggests that the ally relationship may be weaker than that of competition, stronger cues being necessary to elicit response change in a sib comparison situation. This suggestion, although not specified in the original hypotheses of the present study, is consonant with the Adlerian theory.

TABLE 5

SEQUENCE EFFECTS FOR COMPETITORS IN ASPIRATION TASK C

(Scores are mean change scores, $N$ per cell = 10)

| Response Measure | Sample | First Comparison Condition | | | | | Second Comparison Condition | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sib "Better" First (A) | | Sib "Worse" First (B) | | $F$ | Sib "Better" Second (A) | | Sib "Worse" Second (B) | | $F$ |
| | | Mean | *SD* | Mean | *SD* | | Mean | *SD* | Mean | *SD* | |
| LE[a] | Center | +1.17 | 3.58 | −.20 | 2.78 | | −.63 | 2.88 | +2.50 | 4.25 | |
| | Normal | +1.17 | 3.87 | −1.17 | 4.54 | | +.03 | 3.76 | +3.27 | 4.77 | |
| | (Between A & B) | | | | | 1.69 | | | | | 7.15** |
| LA[b] | Center | +3.93 | 7.56 | −1.50 | 9.53 | | +2.10 | 3.22 | +2.30 | 5.15 | |
| | Normal | +6.17 | 9.69 | +.20 | 4.49 | | +.03 | 6.37 | +4.77 | 9.26 | |
| | (Between A & B) | | | | | 5.55* | | | | | 4.28* |

* $p < .05$.

** $p < .01$.

[a] When Center and Normal scores were combined, $F$ for interaction of Better-Worse × First-Second Presentation = 7.59 (for 1,72 *df*, $p < .01$).

[b] When Center and Normal scores were combined, $F$ for interaction of Better-Worse × First-Second Presentation = 5.64 ($p < .05$). Although variances were heterogeneous in the $F$ test for interaction, a $t$ test which compared Better First-Worse Second scores with Better Second-Worse First scores (Center and Normal combined) upheld the significant interaction even though variances were homogeneous for the $t$ test.

TABLE 6

MEAN CHANGE SCORES IN LA AND LE ON ASPIRATION TASK C AS A FUNCTION OF TRIAL SEQUENCE[a]

| Response Measure | Group | Second Set of Three Trials Minus First Three Trials | | Third Set of Three Trials Minus First Three Trials | | $F$ |
|---|---|---|---|---|---|---|
| | | Mean | *SD* | Mean | *SD* | |
| LA | Competitors | +2.20 | 8.64 | +2.30 | 6.61 | |
| | Allies | −1.98 | 6.93 | −2.04 | 8.92 | |
| | Standardization | −2.37 | 8.24 | −3.89 | 10.25 | |
| | Between Competitors & Standardization | | | | | 8.39** |
| LE | Competitors | +.24 | 3.81 | +1.29 | 4.30 | |
| | Allies | −.84 | 7.00 | −1.69 | 9.50 | |
| | Standardization | −2.28 | 11.06 | −3.29 | 11.25 | |
| | Between Competitors & Standardization | | | | | 5.16* |

* $p < .05$. Although variances are heterogeneous, the similarity of trend between LA and LE measures suggests the generality of the Competitor-Standardization difference.

** $p < .01$.

[a] For the experimental *S*s, Better and Worse conditions and Center and Normal *S*s are combined for each cell. $N$ per cell thus = 40 for each of the experimental *S*s. $N$ for each cell of the Standardization *S*s = 20.

TABLE 7
MEAN SCORES ON FIRST THREE (NEUTRAL) TEST TRIALS IN ASPIRATION TASK C
($N$ per cell = 20)

| Group | Performance | | LE | | LA | |
|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| Competitors | | | | | | |
| Normal | 24.30 | 4.88 | 27.97 | 9.30 | 32.75 | 9.46 |
| Center | 25.15 | 4.08 | 26.63 | 5.25 | 34.54 | 7.53 |
| Allies | | | | | | |
| Normal | 25.27 | 5.69 | 28.55 | 8.08 | 33.42 | 8.98 |
| Center | 23.73 | 4.03 | 26.65 | 4.98 | 35.00 | 9.80 |
| Standardization | 22.25 | 4.42 | 29.12 | 6.84 | 34.96 | 11.58 |

Since Allies and Competitors differ in their LA and LE once the sib comparisons are presented to *S*, it could be argued that the groups simply differ in their general aspiration behavior, the sib comparisons being of incidental importance. Inspection of Table 7, however, reveals that Competitors and Allies are comparable in LA, LE, and performance on the neutral trials (and are similar to the standardization *S*s), thus ruling out any basic difference between the two groups in general goal-setting behavior.

## DISCUSSION

The data failed to confirm the original hypotheses of the study. It will be recalled that a revised formulation, based on the original Adler-Dreikurs conceptualization, was suggested as a result of *post hoc* analysis of the data and task dimensions of the present study. This formulation considered ego-involvement and status-effect as crucial variables in an experimental test of the original hypotheses. In the present investigation, the attempt was made to limit all status manipulation to merely providing *S* with information of the sib's performance, all other sources of ego-involvement and status-effect being minimized.[6] However, on the basis of the two variables of ego-involvement and status-effect, it is possible to distinguish Aspiration Task A from Aspiration Task C: Aspiration Task A only elicits ego-involvement and not status-effect since the information given to *S* compares the sib to a norm and not to *S*'s own performance; in Aspiration Task C, in contrast, the attempt is made to evoke status-effect by telling *S* the sib did better or worse in comparison to *S*'s own performance, the comparisons either deflating or inflating *S*'s status. In addition, Aspiration Task C can be assumed to be more ego-involving, if only because *S*s actually entered the task.

The revised formulation takes into account this distinction between the two tasks. It postulates that under conditions of moderate ego-involvement and status-effect (as in Aspiration Task C) Competitors in their goal-setting behave in a manner analogous to rivalry, while Allies are relatively unaffected by the sib comparisons, paying heed primarily to their own rather than the sib's performance. The latter is taken as indicating a weaker relationship for alliance than competition. The revised formulation further postulates that under low ego-involvement and status-effect (as in Aspiration Task A) the two groups do not differ significantly, although both groups may use information of the sib's performance as a cue for their own strivings if they do not have their own performance available as a cue. The fact that Verbal Learning Task B failed to show significant differences between Allies and Competitors is interpreted as in line with the lack of significant effects

[6] Although the Adlerian theory of sibling relations describes the relationships of Alliance and Competition as developing as a function of the child's striving for status (4), it was not recognized until the completion of the present study that manipulation of the status dimensions was an essential component to testing the theory. Thus, for experimental rigor, the elicitation of status variables was explicitly limited to merely providing *S* with information about the sib's performance, so that censure and praise, emphasis of success and failure, etc., were purposely minimized so as not to confound the effects of the sib comparisons. Moreover, neutral tasks were chosen both for the purpose of *not* inducing strong success or failure experiences other than those associated with the sib comparisons, and also so that prior experience with a task would not confound the results.

found in the performance measure on Aspiration Task C. One explanation for the lack of significant effects in the performance measures is that goal-setting possibly taps motivational variables more directly than do performance measures, so that under conditions of moderate ego-involvement and status-effect (as in Tasks B and C) performance measures fail to distinguish between the behavior of the two groups. Under high degree of ego-involvement and status-effect, however, significant differences in performance should be obtained, as stated in the original hypotheses. These hypotheses, drawn from the real-life experience of the child, which presumably contains a high amount of ego-involvement and status-effect, cannot be said to have been tested in the present study.

The revised formulation awaits verification from future studies in which ego-involvement and status-effect are systematically manipulated. The roles of various ego-involving instructions and tasks, as well as of status-effect evoked by censure-praise, failure-success defined by means other than the sib comparisons, etc., can then be judged in their importance to the response changes of the two groups. Allies and Competitors who are compared to a stranger introduced for the purpose of the experiment might also profitably be included in further research. According to Adlerian theory, Allies and Competitors would not differ from one another when compared to a stranger as when compared to the sib, for it is postulated that Allies and Competitors do not intrinsically differ in fundamental personality patterns but that they differ in behavior *only* in response to the sib with whom the particular relationship is established.[7] Data from the present study indicate that the two groups were actually closely comparable,[8] suggesting that the change in goal-setting noted in Aspiration Task C can be attributed with fair certainty to the fact that the *sib* was the compared object.

In conclusion, it must be conceded that the present study leaves several unanswered questions. But the fact that siblings, selected solely on the basis of their likeness or oppositeness of personality traits, differ significantly in goal-setting, when the sib's performance is used as the only source of status-effect, suggests the fruitfulness of further experimental investigation of the Adlerian theory of sibling relations.

## SUMMARY

The present study set out to test hypotheses derived from the Adlerian theory of sibling relations. Two duplicate experiments were performed, 20 sibling pairs being drawn from Child Guidance Centers and 20 pairs from a "normal" population. Each group contained 10 pairs of Allies and 10 pairs of Competitors, the operational definition of alliance being sameness of traits of the siblings and competition being oppositeness of traits. Diagnosis of the sibling relations was based on interview material obtained from the mother of each pair of sibs. Each *S* was tested singly and was given two aspiration tasks and one verbal learning task. On each task, *S* was given predetermined information about the sib's performance. Although the hypotheses tested by the study were not upheld, Allies and Competitors were found to differ significantly in their goal-setting when the sib's performance was compared to *S*'s own performance. A formulation that revised the original hypotheses of the study was suggested, with the variables of ego-involvement and status-effect being considered crucial to an adequate test of the original hypotheses. Although the present study leaves several unanswered questions,

[7] That is to say that although a given *pair* of Allies will be alike in personality and a given pair of Competitors opposite in personality, a random *group* of Allies will show the same range in personality traits as a random group of Competitors. Thus, out of four Allies, one pair may be striving and dominant and another pair retiring and submissive, while in each of two pairs of Competitors one sib may be striving and dominant and the other sib retiring and submissive. The result will be that two out of the four Allies will be like two out of the four Competitors, and the other two Allies will be like the other two Competitors. In this example, it would be reasonable to expect that the four dominant *S*s will act differently toward a stranger than the four submissive *S*s. Such data would therefore contrast to that of the present study in which the *sib* was the compared object, in that responses to a stranger would not differentiate between the behavior of Allies and Competitors.

[8] Evidence for the comparability of the two groups derives from the following: (*a*) the two groups had comparable scores on the neutral trials of Aspiration Task C; (*b*) the interview data revealed the personalities of the two groups to be quite similar; (*c*) many of the Allies had sibs with whom they had a competitor relationship while several of the Competitors had sibs with whom they had an alliance relationship.

the significant results that were obtained suggest the fruitfulness of further experimental investigation of the Adlerian theory of sibling relations.

## REFERENCES


1. ADLER, A. *The science of living*. New York: Greenberg, 1929.
2. ADLER, A. *The education of children*. London: Allen & Unwin, 1930.
3. CHAPMAN, L. J. An investigation of loss of set in the conceptual performance of schizophrenics. Unpublished doctoral dissertation, Northwestern Univer., 1954.
4. DREIKURS, R. *Fundamentals of Adlerian psychology*. New York: Greenberg, 1950.
5. DREIKURS, R. The psychological interview in medicine. *Amer. J. indiv. Psychol.*, 1952–53, **10**, 99–122.
6. FERGUSON, EVA D. The effect of sibling competition and alliance on level of aspiration, expectation, and performance. Unpublished doctoral dissertation, Northwestern Univer., 1956.
7. HOLT, R. R. Effects of ego-involvement upon levels of aspiration. *Psychiatry*, 1945, **8**, 299–317.
8. IVERSON, M. A., & REUDER, M. E. Ego involvement as an experimental variable. *Psychol. Rep.*, 1956, **2**, 147–181.
9. WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.


Received April 8, 1957.

# EFFECTS OF PUBLIC AND PRIVATE FAILURE ON SELF-EVALUATION[1]

EZRA STOTLAND[2] AND ALVIN ZANDER

*Research Center for Group Dynamics, University of Michigan*

A PERSON who has failed may as a result lower his evaluation of himself, under certain conditions, yet under other conditions he may not pay much attention to the failure. Among the circumstances that may affect the nature of his reaction are his opinions about how others evaluate his performance, and these opinions are determined by whether the others know about the failure.

It is evident that a person who has failed tends to become sensitive to the opinions of others. This has been shown in several empirical studies (2, 3, 4). It is known, too, that a person's opinion about himself is affected by the evaluations of his performance that he attributes to others (4). If the failer assumes that others think poorly of his performance, his feelings about himself are likely to depend upon how much he attributes validity to the others' judgments about him.

In this study we seek to answer three questions. Two of these are: Will a person's evaluation of himself when he fails in complete privacy be different from his evaluation when he is aware that another knows about the quality of his performance? Will a person's evaluation of himself in the presence of an incompetent judge of his ability be different from his rating when the other person is a competent judge?

In attempting to answer these questions, we take it as given that a person at all times wishes to evaluate himself as highly as it is possible for him to do, within the limits set by the facts as he sees them. When an individual fails on a task and clearly knows that he has done so, he cannot deny his failure to himself without grossly distorting reality. Thus, he is not likely to deny failure in such an instance. But where different degrees of failure are possible and the degree of his failure is uncertain in his own mind, he can deny that he has failed at all, or minimize the degree of the failure.

[1] The work on this study was financed by a grant-in-aid (M-701 (C)) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

[2] Now at University of Washington.

A person who has failed has no reason to believe that other individuals who are aware of his failure, especially strangers, minimize the extent of that failure. Thus, we assume that a person who has failed believes that others will tend to evaluate his performance as worse than he would like to believe it was.

If a failer, then, conforms to the evaluations he assumes another has of him, he will evaluate his performance low. What will lead him to accept a stranger's negative evaluations? First, it is necessary that the other know about the failure, since unless the other is aware of the poor performance, the failer has few constraints against minimizing his failure to himself. In addition, for him to conform to the other's negative evaluation, the other must generate social pressures on the one who has failed, pressures which are strong enough to overcome the failer's tendency to set his self-evaluation as high as possible. It is assumed that stronger social pressures are exerted by persons who are experts than by those who are nonexperts.

Therefore, it is first hypothesized that, 1(*a*) a person who had failed on a task evaluates the degree of his failure as being greater when his failure is observed by an expert than when it is not observed by an expert; 1(*b*) the observation or nonobservation by a nonexpert, however, does not generate variations in self-evaluations, but 1(*c*) the person evaluates his performance as poorer when his failure is observed by an expert than when it is observed by a nonexpert.

Thus far, the individual has been considered as evaluating only the quality of his performance on a task, as good or poor. He may in addition make evaluations of personal abilities that were involved in the failure. Because his evaluation of his abilities is likely to be affected by the evaluation of his performance, his rating of the abilities will probably be similar to his evaluation of his performance. The three hypotheses described above, therefore, apply equally well to evaluations of abilities involved in the person's performance on the task.

We conceive abilities as being either closely related or less closely related to the ones under test. Further, we assume that the abilities most closely related to those being employed at the time of his failure will be affected by the level of that evaluation, while those less closely related will be less affected by the evaluation. We raise then a third question: How much of the person's self is included in his evaluations of himself after an experience of failure? A second hypothesis is proposed: Where a person gives a low evaluation of his performance, because of observation of his failure by an expert, the evaluation of closely related abilities will also be low while the evaluation of less closely related abilities will not be as low.

## METHOD

These hypotheses were tested in a laboratory experiment. The *S*s were 126 male undergraduate volunteers. As each arrived in the laboratory he was privately given a puzzle which was described as a test of visual motor coordination, with completion of the puzzle as the criterion of success. Before the *S*s worked on it, *E* introduced himself to half of them as an expert on such tests; to the other half as a nonexpert. The *S*s worked in privacy on the task, which was impossible to complete. Then, in half the cases, *E* entered the *S*s' rooms and told them that he observed they had failed (public condition); in the other half, he asked the *S*s, while outside their rooms, to put the test material away without leaving any indication of how well they had done (private condition). All the participants then privately answered a questionnaire in which they evaluated their performances on the task, their visual motor coordination, and their abilities in areas varying in relevance to the task. The separate phases of the experiment are now described.

### *Experimental Task*

The *S*s arrived individually in the laboratory and *E* escorted them immediately to private cubicles where they remained for the rest of the experiment. Each *S* was told he was to work on a puzzle to help him discover things about himself so that he could perform better in the later or "real" parts of the experiment, although he was not told what these "real" parts would be. He was told that working on these tasks would teach him something useful about himself. The purposes of these instructions were to arouse the *S*s' interest in self-evaluation, to motivate them to work on the puzzle, and to interest them in answering the postexperimental questionnaire for their own benefit, not as a communication to a research worker. In this way, it was hoped to obtain more honest and valid answers.

The first task was a puzzle which was presented as "an excellent test of visual-motor coordination." Visual-motor coordination was defined for the *S* as the "ability to imagine how certain shapes and forms fit together and then to take this arrangement and put it down in actuality." Thus, the test was made relevant to certain specified abilities. It was assumed that visual-motor coordination was an aspect of themselves on which these persons did not have fixed evaluations and was, therefore, subject to experimental influence. The puzzle consisted of 12 irregular cardboard pieces with which *S* was to fill the space inside an irregular outline traced on a board. He was allotted ten minutes. The criterion of complete success was said to be full completion of the puzzle. *S* was told that degrees of failure were determined by the discrepancy between the achieved arrangement and the one possible way of completing the puzzle. The number of pieces employed in the design were described as irrelevant to his degree of failure. In this way the objective degree of failure was made ambiguous for the *S*, allowing social factors more influence on his evaluation of his performance.

Before *S* was allowed to work on the puzzle, *E* showed him his second task, the postexperimental questionnaire. This was left in the cubicle for him to fill out after the time allotted for work on the puzzle had been spent. The private nature of his answers was emphasized again by telling him that answers were not of interest to the *E* and by repeating some of the earlier instructions.

### *Experiness Induction*

Before *S* worked on the puzzle, *E* introduced himself in half the cases as Dr. X who had much experience in administering and constructing tests of visual-motor coordination (expert condition). In the other half, *E* introduced himself as a substitute for the regular experimenter and as one who knew little about these tests (nonexpert condition).

### *Public-Private Induction*

The *S* was asked to work privately on the task. The *E* announced that the allotted time had elapsed while standing outside the *S*'s cubicle. In the public condition, *E* entered the cubicle and said impersonally, "I see that you were not able to complete the task. I'm afraid this would have to be called a failure." He then instructed *S* to put the pieces of the puzzle back into the envelope in which he had found them and to go on to the questionnaire. In the private condition, *E* did not enter the cubicle but instead instructed *S* to put the pieces back into the envelope in which he had found them and to go on with the questionnaire. It should be remembered that in both conditions the *S*s knew they had failed to some degree without *E*'s saying so, since they had been told that the criterion of full success was completion of the puzzle.

Thus a four-cell, 2 × 2 design was employed with the following conditions: public-expert, public-nonexpert, private-expert, and private-nonexpert.

### *Post-experimental Questionnaire*

To develop a series of items about self-evaluation that had various degrees of similarity to the abilities involved in the experimental task, a list was obtained (from work as yet unpublished) of the traits that are used most frequently by male undergraduates in describing themselves. This list was ranked a priori by the investigators in respect to their degree of similarity to the abilities needed for the experimental task. To modify this a priori ranking so that it corresponded to the ranking that would be made by an average male undergraduate, a modification of a method used by Coombs (1) was employed, using male undergraduate *S*s other than those in the experiment.

The first item on the questionnaire asked *S* to rate his success on an 11-point scale, if he felt that he had succeeded on the puzzle. The second question was a parallel one for failure, the *S* answering one but not both since all of them had failed. In this way his evaluation of his *performance on the puzzle* was measured. He next made an evaluation of his *visual-motor coordination* (hereafter designated as VMC) and of a series of 13 *abilities* presented in order of decreasing similarity to the abilities required in the experimental task. In the analysis of the results, this 13-item scale was divided into three groups of 4, 4, and 5 items each and the mean evaluation attributed to each group was obtained. These means are referred to hereafter as M1, M2, and M3, the subscripts designating the order of decreasing similarity to the abilities required in the experimental task.

The ratings given to VMC and M1 are assumed to indicate evaluations of abilities that are closely related to the ones used in performing the task, while evaluations of M2 and M3 are assumed to concern areas less related to the task-required abilities. In addition to these 13 items, a final question requested that the *S* evaluate himself as a "complete person."

Finally, they were asked to rate the competence of the *E* on an 11-point scale. This question was inserted in order to determine the extent to which *S*s believed *E* to be an expert or a nonexpert in these two conditions.

## RESULTS

To determine whether the *E* was perceived by the *S*s as being either expert or nonexpert, the mean ratings of the competence they attributed to *E* in the post questionnaires were computed. The mean for the expert condition was 7.19, and for the nonexpert condition 5.56, significantly different at the .001 level of probability ($t = 3.62$). It is evident that the *S*s on the whole accepted *E*'s descriptions of his expertness in the separate conditions; but those who were told that the *E* was not an expert attributed more competence to him than might be expected.

TABLE 1
MEAN SELF-EVALUATIONS FOR VARIOUS EXPERIMENTAL CONDITIONS

| Item Evaluated | Experimental Conditions | | | | $p$ value of diffs.* |
|---|---|---|---|---|---|
| | Public Expert $N = 25$ Col. a | Public Non-expert $N = 39$ Col. b | Private Expert $N = 34$ Col. c | Private Non-expert $N = 28$ Col. d | |
| Performance on puzzle | 2.58 | 1.52 | 2.75 | 4.20 | b–d .001 |
| Visual motor coordination | 3.79 | 4.10 | 4.78 | 4.31 | a–c .06 |
| Abilities M1 | 5.07 | 5.02 | 5.76 | 5.37 | |
| Abilities M2 | 6.88 | 6.48 | 6.76 | 6.92 | |
| Abilities M3 | 6.11 | 6.24 | 6.44 | 6.22 | |
| Complete person | 6.17 | 6.48 | 6.71 | 6.34 | |

* By two-tail $t$ tests.

The mean self-evaluations for their performances on the puzzle and for their separate abilities are contained in Table 1. These data provide no support for hypotheses 1(*a*), 1(*b*), or 1(*c*) in respect to evaluation of performance on the puzzle. In fact, contrary to what was anticipated in hypothesis 1(*b*), the *S*s who were observed by a nonexpert evaluated their performances more poorly than those who were not observed by him, the difference being significant at the .001 level.

The evaluation given to their visual-motor coordination (VMC), is lower in the public-expert condition than in the private-expert condition ($p$ of diff. = .06), as would be expected by hypothesis 1(*a*), and observation or nonobservation by a nonexpert generated no difference in evaluation of VMC as would be expected by hypothesis 1(*b*) These data cannot be accepted as support for hypotheses 1(*a*) nor for 1(*b*), however, since it is apparent that some of these *S*s did not behave as was predicted in evaluating their performances, and the ratings of VMC depend upon a generalization of the evaluation of their performances to their visual-motor coordination as an ability. The evaluations assigned to M1 abilities were approximately the same in all conditions. It is apparent, then, that the data in Table 1 provide no convincing corroboration for hypotheses 1(*a*), 1(*b*), or 1(*c*).

In order to control on the effects of the evaluations given to their performances on the puzzle in determining the evaluations of VMC and M1 abilities, the total sample was divided into two groups at the median of their evaluations of their performances on the puzzle.

TABLE 2

MEAN SELF-EVALUATIONS FOR VARIOUS EXPERIMENTAL CONDITIONS FOR SUBJECTS WITH HIGH AND LOW EVALUATIONS OF PERFORMACE ON PUZZLE

| Item Evaluated | Experimental Conditions: Public Expert $N = 11$ Col. a | Public Nonexpert $N = 12$ Col. b | Private Expert $N = 18$ Col. c | Private Nonexpert $N = 16$ Col. d | Total Col. e | $p$ value of diffs.[a] |
|---|---|---|---|---|---|---|
| Part A—Subjects with high evaluation of performance on puzzle | | | | | | |
| Visual motor coordination | 3.64 | 5.15 | 5.17 | 5.00 | 4.83 | a–c .03<br>a–b .03 |
| Abilities M1 | 5.03 | 5.87 | 5.97 | 5.83 | 5.73 | a–b .05<br>a–c .05 |
| Abilities M2 | 6.60 | 7.03 | 6.87 | 6.86 | 6.85 | |
| Abilities M3 | 6.15 | 6.68 | 6.63 | 6.00 | 6.26 | |
| Complete person | 6.18 | 7.09 | 6.72 | 6.44 | 6.61 | |
| | $N = 13$ | $N = 19$ | $N = 9$ | $N = 9$ | | |
| Part B—Subjects with low evaluation of performance on puzzle | | | | | | |
| Visual motor coordination | 3.92 | 3.42 | 3.89 | 3.00 | 3.56 | |
| Abilities M1 | 5.15 | 4.49 | 4.80 | 4.28 | 4.68 | |
| Abilities M2 | 7.13 | 6.14 | 6.44 | 6.44 | 6.51 | a–b .001 |
| Abilities M3 | 6.52 | 5.97 | 6.02 | 5.96 | 6.12 | |
| Complete person | 6.08 | 6.11 | 6.56 | 5.22 | 6.20 | |

Note.—Value of $p$ for differences between comparable cells in Parts A and B of Table 2: VMC, Col. b, .01; VMC, Col. d, .05; VMC, Col. e, .001; M1, Col. b, .001; M1, Col. c, .05; M1, Col. d, .03; M1, Col. e, .001.

[a] By two-tail $t$ tests.

Then, the means of the evaluations given to their separate abilities were calculated. These are reported in Table 2.

It can be seen there that those *S*s who minimized their failures and gave better evaluations to their performances (Part A), tended in the large majority of cases to give higher evaluations to their abilities than did *S*s (Part B) who rated their performances as poorer. There is a strong tendency, it is clear, for persons who think better of their performances to also think better of their abilities. This tendency is evident in every experimental condition, except the public-expert one.

In the public-expert condition, the *S*s who had rated their performances high evaluated their VMC and M1 abilities as low as did the *S*s who had rated their performances low. We return later to a discussion of this discrepancy in the evaluations of performances and abilities in the public-expert condition.

The low evaluations of the VMC and M1 abilities by *S*s in the public-expert condition, but not in the other conditions, provide data relevant to the first hypothesis.

The (*a*) part of Hypothesis 1 asserts that persons make lower evaluations of themselves when their failures are observed by an expert than when not observed. It is clear that VMC and M1 abilities are rated significantly lower by *S*s who were in the public-expert conditions than by those who were in the private-expert condition. This is true, however, only for those who evaluated highly their performances on the puzzle.

The (*b*) part of Hypothesis 1 states that where the *E* is a nonexpert, persons do not evaluate themselves differently when their failure is observed as compared to when it is not observed. The hypothesis is supported both by those who rated their performance high, and by those who rated their performance low.

The (*c*) part of Hypothesis 1 says that observation of the *S*s' failures by an expert generates lower self-evaluations than does observation by a nonexpert. It is evident that this hypothesis is supported by the responses of those *S*s who gave themselves high evaluations in their performance on the puzzle, but is not supported by the *S*s who had low evaluations of their performance on the puzzle.

It is clear, then, that two of the three parts of the first hypothesis are supported by ratings given to VMC and M1 abilities only where the *S*s felt that, or stated that, they had performed well on the puzzle.

Among the *S*s who made low evaluations of their performances (Part B of Table 2) there are no significant differences in the evaluations of VMC and M1 abilities when the means of the four experimental conditions are compared. Perhaps these *S*s evaluated their performances on the puzzle, and their VMC and M1 abilities, so low that the pressure on them, stemming from the *E* to lower their evaluations of themselves could cause no further drop in self-evaluations regardless of the strength of these pressures. Where the *S*s had made high evaluations of their performances on the puzzle, the height of these evaluations apparently provided "room" for the social pressures effectively to lower evaluations of closely related abilities.

Hypothesis 2 states that where a person gives a low evaluation of his performance because his failure has been observed by an expert, the evaluation of closely related abilities will also be low, while the evaluations of less closely related abilities will not be as low. The results in Table 1 are inconclusive in respect to this hypothesis, although they are in the predicted direction.

In Part A of Table 2, as we have seen, the abilities which are most closely related to performance on the puzzle (VMC and M1) were evaluated significantly lower where the *E* was an expert and observed the *S*s' failures and where the *S*s had placed a high evaluation upon their performances on the puzzle. It is evident that abilities M2, M3, and the "complete person" were not, however, rated lower where *E* was an expert and observed the *S*s' failures. These results support Hypothesis 2. In Part B of Table 2, *S*s also evaluated their M2 abilities lower, for some unexplainable reason, when their failures had been observed by a nonexpert rather than an expert.

It is also relevant to note that participants who gave low evaluations to their performances on the puzzle, compared to those who gave high evaluation to their performances, attributed significantly lower evaluations to their VMC and M1 abilities when the responses of *S*s from all conditions were combined (Col. e in Table 2). There were no significant differences, however, in the ratings given to M2, M3, or the "complete person" (all conditions combined) when the responses of persons with high evaluations of their performance were compared with those who had low evaluations of their performances. These results appear again to indicate that low evaluation of one's original performance affect evaluations of the most closely related abilities more than they affect evaluations of those abilities which are less closely related.

## DISCUSSION

The results suggest a modification of the assumptions and hypotheses, but a modification that is consistent with the rest of the theorizing. In the statement of the theory at the outset of this paper, it had been assumed that individuals would tend to evaluate themselves as highly as possible within the limits of reality, but that pressures toward more objective evaluations from other persons would cause the *S*s to depress their evaluations of their performances and abilities. The data suggest, however, that some persons who fail on a task make such low evaluations of their performances on that task, and of their abilities related to their performances, that social pressures could not lead to any further depression in their self-evaluations. Thus, it was found that *S*s who made low self-evaluations of their performances evaluated their closely related abilities uniformly low, irrespective of whether they were being observed by the *E* and whether he was an expert.

On the other hand, the high evaluations that some *S*s made of their performances can be taken as a sign that they were motivated to make high evaluations of themselves, more than were those who made low evaluations of their performances. For the highly motivated *S*s, variations in the degree of publicness of the *S*s' failures and in the expertness of the *E* had the predicted consequences on their evaluations of abilities. When they were observed by an expert, they conformed to his presumably stronger influences and to his evident low evaluations, and thus evaluated their abilities as poorly as did the *S*s with low evaluations of their performances. In line with the original hypotheses, those *S*s with high evaluations of their performances rated lower abilities that were more closely related to the performances on the puzzle when observed by an expert than when not observed by him, or when observed by a nonexpert. Also, in keeping with the hypotheses, their evaluations of

abilities were not different when they were observed by a nonexpert than when not observed by him. Where the assumption that a person will try to keep his self-evaluation as high as possible is shown by the data to be present and operating, then, the experimental hypotheses were supported in regard to the *S*s' evaluations of their relevant abilities. Where this assumption was not fulfilled, the hypotheses were not supported. Instead of assuming, as was done in the opening paragraphs of this paper, that there is a tendency in all persons at all times to evaluate themselves as highly as possible, we must assume that this tendency is present in some persons. Where it is present, the hypotheses proposed for this study are best fulfilled.

This modification of the assumptions and hypotheses raises the question as to why some *S*s attributed high evaluations to their performance on the puzzle, while others did not. In particular, the question is raised in reference to the *S*s in the public-expert condition in Table 2. No conclusive answer to this question is available, although it is likely that some of the *S*s felt more need to look good in the eyes of the *E* even when they made their ratings privately. It is possible that the high self-ratings in the public expert condition may have been due to resistance to the stronger pressures exerted by the *E* in that condition. A complete answer probably requires some analysis of the motivations of the *S*s when rating themselves.

These results point to the need for a theoretical and experimental refinement of the meaning of "expertness" for interpersonal relationships. It is not clear from the present experiment whether it is the prestige of the expert, the power of the expert, or the objectivity he assumes, which generates the obtained differences in the evaluations of the *S*s in the separate conditions. A crucial test would be to obtain self-evaluations from *S*s who were observed by a person whose prestige was varied independently of variations in his apparent ability to be competent in judging the *S*s' performances.

The results support the second hypothesis of the study, namely, that the differences among the conditions are limited to evaluations of abilities most closely related to the abilities required for performance of the task. Only one significant difference was found among the conditions in the evaluations of other areas of the self, in all probability a result without psychological significance.

## SUMMARY

In this study we have examined how variations in the strength of social pressures, acting upon a person in the direction of lowering his self-evaluation, affect the evaluations the individual assigns to the quality of his performmance, and to his more or less closely relevant abilities, after he has failed on a task. Social pressures were assumed to be stronger on the person when his performance was viewed by an expert than when it was not observed, or when it was observed by a nonexpert. Social pressures were assumed to restrain a person from assigning a high evaluation to the quality of his behavior or abilities.

The subjects were individually and privately each asked to solve an impossible puzzle which was described as a test of visual-motor coordination. In half of the cases their level of performance was observed by the experimenter; in the other half it was not observed. Within each of these conditions the experimenter introduced himself to half the subjects as an expert and to the other half as a nonexpert. After the subjects had failed, they rated their performances and abilities. The abilities had varying degrees of similarity to those under test in the puzzle.

It was found that:

1. *S*s who made high evaluations of their performances on the puzzle, evaluated abilities that were involved in their performances lower when they had been observed by an expert than when they had not been observed by an expert, or when they had been observed by a nonexpert.

2. *S*s who made low evaluations of their performances did not evaluate these abilities differently under different experimental conditions.

3. No consistent consequences from the publicness or privacy of failure, nor from the expertness of the experimenter were found when the results from the entire population of subjects was examined in respect to the

evaluations of the quality of their performances and their abilities.

4. The differences in the effects of the experimental conditions were limited to those abilities which were most closely related to the abilities required by the task; evaluations of other parts of the self were less affected.

It was concluded that strong social pressures acting on a person toward lower evaluations cause more lowering in his evaluation of his abilities when the person is strongly motivated to keep his evaluations high, than when he has little motivation to keep these evaluations high.

## REFERENCES

1. COOMBS, C. H. A method for the study of interstimulus similarity. *Psychometrika*, 1954, **19**, 183–194.
2. JACKSON, J., & SALTZSTEIN, H. D., *Group membership and conformity processes*. Res. Center for Group Dynamics, Ann Arbor, 1956.
3. RASSMUSSEN, G., & ZANDER, A. Group membership and self-evaluation. *Hum. Relat.*, 1954, **7**, 239–251.
4. STOTLAND, E., THORLEY, S., THOMAS, E., COHEN, A. R., & ZANDER, A. The effects of group expectations and self-esteem upon self-evaluation. *J. abnorm. soc. Psychol.*, 1957, **54**, 55–63.

Received April 23, 1957.

abilities were not different when they were observed by a nonexpert than when not observed by him. Where the assumption that a person will try to keep his self-evaluation as high as possible is shown by the data to be present and operating, then, the experimental hypotheses were supported in regard to the *S*s' evaluations of their relevant abilities. Where this assumption was not fulfilled, the hypotheses were not supported. Instead of assuming, as was done in the opening paragraphs of this paper, that there is a tendency in all persons at all times to evaluate themselves as highly as possible, we must assume that this tendency is present in some persons. Where it is present, the hypotheses proposed for this study are best fulfilled.

This modification of the assumptions and hypotheses raises the question as to why some *S*s attributed high evaluations to their performance on the puzzle, while others did not. In particular, the question is raised in reference to the *S*s in the public-expert condition in Table 2. No conclusive answer to this question is available, although it is likely that some of the *S*s felt more need to look good in the eyes of the *E* even when they made their ratings privately. It is possible that the high self-ratings in the public expert condition may have been due to resistance to the stronger pressures exerted by the *E* in that condition. A complete answer probably requires some analysis of the motivations of the *S*s when rating themselves.

These results point to the need for a theoretical and experimental refinement of the meaning of "expertness" for interpersonal relationships. It is not clear from the present experiment whether it is the prestige of the expert, the power of the expert, or the objectivity he assumes, which generates the obtained differences in the evaluations of the *S*s in the separate conditions. A crucial test would be to obtain self-evaluations from *S*s who were observed by a person whose prestige was varied independently of variations in his apparent ability to be competent in judging the *S*s' performances.

The results support the second hypothesis of the study, namely, that the differences among the conditions are limited to evaluations of abilities most closely related to the abilities required for performance of the task. Only one significant difference was found among the conditions in the evaluations of other areas of the self, in all probability a result without psychological significance.

## SUMMARY

In this study we have examined how variations in the strength of social pressures, acting upon a person in the direction of lowering his self-evaluation, affect the evaluations the individual assigns to the quality of his performance, and to his more or less closely relevant abilities, after he has failed on a task. Social pressures were assumed to be stronger on the person when his performance was viewed by an expert than when it was not observed, or when it was observed by a nonexpert. Social pressures were assumed to restrain a person from assigning a high evaluation to the quality of his behavior or abilities.

The subjects were individually and privately each asked to solve an impossible puzzle which was described as a test of visual-motor coordination. In half of the cases their level of performance was observed by the experimenter; in the other half it was not observed. Within each of these conditions the experimenter introduced himself to half the subjects as an expert and to the other half as a nonexpert. After the subjects had failed, they rated their performances and abilities. The abilities had varying degrees of similarity to those under test in the puzzle.

It was found that:

1. *S*s who made high evaluations of their performances on the puzzle, evaluated abilities that were involved in their performances lower when they had been observed by an expert than when they had not been observed by an expert, or when they had been observed by a nonexpert.

2. *S*s who made low evaluations of their performances did not evaluate these abilities differently under different experimental conditions.

3. No consistent consequences from the publicness or privacy of failure, nor from the expertness of the experimenter were found when the results from the entire population of subjects was examined in respect to the

evaluations of the quality of their performances and their abilities.

4. The differences in the effects of the experimental conditions were limited to those abilities which were most closely related to the abilities required by the task; evaluations of other parts of the self were less affected.

It was concluded that strong social pressures acting on a person toward lower evaluations cause more lowering in his evaluation of his abilities when the person is strongly motivated to keep his evaluations high, than when he has little motivation to keep these evaluations high.

REFERENCES

1. COOMBS, C. H. A method for the study of interstimulus similarity. *Psychometrika*, 1954, **19**, 183–194.
2. JACKSON, J., & SALTZSTEIN, H. D., *Group membership and conformity processes*. Res. Center for Group Dynamics, Ann Arbor, 1956.
3. RASSMUSSEN, G., & ZANDER, A. Group membership and self-evaluation. *Hum. Relat.*, 1954, **7**, 239–251.
4. STOTLAND, E., THORLEY, S., THOMAS, E., COHEN, A. R., & ZANDER, A. The effects of group expectations and self-esteem upon self-evaluation. *J. abnorm. soc. Psychol.*, 1957, **54**, 55–63.

Received April 23, 1957.

# DISTANCE AS A DETERMINANT OF INFLUENCE IN A TWO-PERSON SERIAL INTERACTION SITUATION

SEYMOUR FISHER AND ARDIE LUBIN[1]

*Walter Reed Army Institute of Research*

THIS paper is concerned with two major aspects of a two-person social influence situation. The first deals with the systematic exploration of the effects of "distance" upon measures of influence, where distance refers to the size of the discrepancy between the subject's (*S*'s) pre-influence judgment and the subsequent judgment originating from an influence source (6). The concept of distance corresponds closely to "opinion deviance" as used by Festinger (2), Schachter (20), Goldberg (8), and others; in Asch's work (e.g., 1), he speaks of "size of discrepancy"; French (7) employs the same phrase in his model. These writers, along with Newcomb (18), have all suggested in one form or another that distance is an important variable in determining how much *S*'s judgment can be altered in a social interaction. The precise relationship, however, seems to be in doubt. Do large distances produce more "influence" than smaller distances? One difficulty in attempting to answer such a question lies in the ambiguity of the term "influence." As noted in a previous paper (6), influence can be measured in at least two ways: (*a*) how much *S* *moved* from his original judgment, and (*b*) how much *S* *conformed* to the source's judgment (i.e., *S*'s movement relative to the distance). So long as distance is held constant, measures of movement and conformity are in a one-to-one relationship; hence, either can be used satisfactorily. When *S*s are exposed to different distances, however, the two measures are no longer directly comparable and yield different results. An *S* who moves 100 units away from his pre-influence judgment under a distance pressure of 200 is behaving quite differently from another *S* who moves 100 units under a distance pressure of 1000; while both *S*s have *moved* equal amounts, the former *S* has *conformed* .50 (i.e., has moved half-way toward the influence source's judgment) as contrasted with the latter *S* who has conformed only .10. One of the purposes of this study was to investigate parametrically the way in which movement and conformity are related to the independent variable of distance.

The second problem under investigation concerns behavior in a *serial interaction situation*, where *S* is exposed to a number of influence attempts on successive, discrete stimuli, each common to the same broad dimension. Whereas most experimental studies seem to focus on the effects of one or more influence trials treated as a single influence attempt (measuring the difference between pre- and post-influence judgments), it is more usual in real life that *S*s are exposed to a set of influence attempts from one source, sometimes on the same issue, sometimes on different issues. In the present study, the concept of "serial interaction" was experimentally simplified by studying *S*s' reactions to two consecutive stimuli. Three questions were being asked here: (*a*) What are the effects of repeated influence attempts on the *same* stimulus? (*b*) What carry-over effects does a series of influence attempts on one stimulus have upon the perception of a second, similar stimulus? (*c*) What effects does a series of influence attempts on one stimulus have upon subsequent influence attempts on a second, similar stimulus?

## METHOD

The data have been obtained from two separate studies, the second essentially a replication of the first. The *S*s were army enlisted men from Fort Meade, Maryland. Both studies required *S*s to judge the number of paratroopers shown in briefly exposed photographs. In the original experiment, 96 *S*s were used in a 3 x 2 x 2 factorial design. The independent variables consisted of three different distances (40, 320, and 5100), two experimenters (*E*s), and two sets of instructions (one emphasizing accuracy, the other emphasizing agreement with partner). A pilot study showed that a significant experimenter-distance interaction would result unless instructions were rigidly standardized; in the study, instructions, *E*s, and interactions were analyzed—none was significant. The second experiment used a total of 192 *S*s, divided into six distance groups (0, 40, 160, 320, 1300, 5100), 32 *S*s per group; two different *E*s ran the study, with standardized instructions emphasizing accuracy.

Two *S*s at a time were brought into the laboratory

[1] Grateful acknowledgment is made to the invaluable technical assistance of Esther Adams, Nicholas Covatta, Peter Kyriazis, Major E. E. Admire, and Hilda Wexler.

and seated side by side in front of *E*, who read the preliminary instructions to them. In brief, these instructions said that *E* was interested in seeing how accurately people could judge the number of troops shown in photos for short periods of time; further, whether people could be more accurate when they were working together and able to talk things over than when they were by themselves and could not discuss their judgments. Each *S* was given a pad of colored paper to write down his judgments, the color of the pad serving as identification of *S*.

### *"Warm-up" Period*

*E* began by exposing a photo for 10 seconds. After the photo was removed, the two *S*s wrote their independent estimates on their pads. The estimates were announced to the *S*s, and (if *S*s disagreed) they were told to "talk it over between yourselves and try to be as accurate as possible" (accuracy instructions). After a few minutes of discussion, each wrote down his second estimate, after which *E* announced the true number, and *S*s were permitted to examine the photo and verify *E*'s statement. This warm-up period was extended over five photos and served a number of important functions: (*a*) *S*s became accustomed to estimating the number of paratroopers, and also became familiar with the procedure of making more than one estimate to the same photo. (*b*) The credibility of the subsequent experimental phase was enhanced. Each time *S* wrote down an estimate during the warm-up, the papers were taken by *E* to an assistant a few feet away (and partially out of sight) who recorded and audibly stamped the estimates by machine. Thus, in the experimental phase which followed, *S*s accepted the routine of having *E* take their estimates to the assistant *before* they were exchanged (permitting the assistant to intercept the original estimates and exchange them for bogus messages).

The *S*s were then told that for the remaining time they were to work alone, with no talking between them; all they could do was exchange their estimates. They were placed in separate, adjacent cubicles where they could not see each other. Each *S* retained the same colored pad he had used in the "warm-up."

### *Experimental Period*

In the account of the procedure which follows, the major measures which were analyzed are indicated in parentheses immediately following the operations to which they refer.

*E* exposed Card A, a photo containing 165 paratroopers, for a 10-second interval, and *S*s wrote down their initial estimates ($A_1$). *E* picked up the estimates to be stamped and recorded, during which time the bogus estimates were written by the assistant on appropriately colored paper. Hence, when the estimates were "exchanged," each *S* was led to believe that he was receiving his partner's estimate, whereas he actually received an estimate which was a fixed distance *greater* than his initial estimate (except, of course, that *S*s in the zero distance group received bogus estimates identical with their $A_1$). *S*s then made their second estimate ($A_2$), the estimates were again "exchanged," and a total of five estimates was made by each *S* to

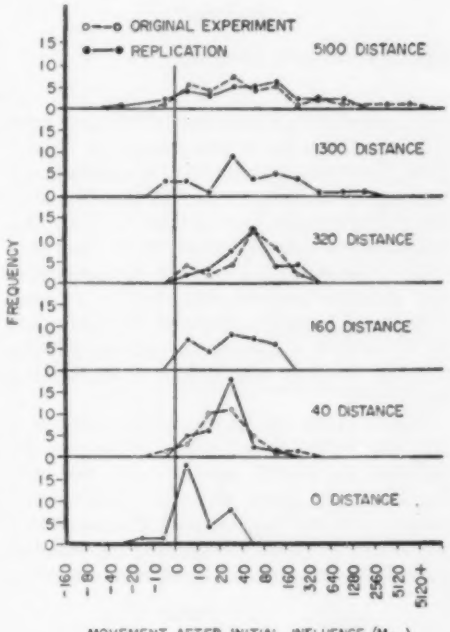


FIG. 1. FREQUENCY DISTRIBUTIONS OF INITIAL MOVEMENT SCORES ON CARD A AS A FUNCTION OF DISTANCE (Note abscissa is scaled geometrically)

Photo A. *The bogus estimates always remained the same* for any *S* throughout the five interchanges, regardless of whether or not *S* changed his estimate at any time. After the fifth estimate, Photo A was exposed again for 10 seconds, and *S*s were asked to make a "final" estimate ($A_u$).[2] $A_u$ and $A_1$ were the only estimates made immediately after seeing Photo A, without an intervening bogus estimate.

Card B, a photo containing approximately 300 paratroopers, was then shown for 10 seconds, and *S*s made their initial estimates ($B_1$) to this photo. The messages were "exchanged," but this time *all S*s, regardless of the distance received on Photo A, *received a bogus estimate 320 greater than* $B_1$. The *S*s made a second estimate ($B_2$), after which Photo B was again exposed for 10 seconds, and *S*s made a "final" estimate ($B_u$). An additional measure, taken only in the replication study, consisted of giving *S*, at the end of the experiment, Photos A and B for a period of three minutes during which time he was asked to count the actual number of paratroopers in the two photos ($A_v$ and $B_v$, respectively).

[2] In the replication, half of the *S*s were not given the $A_u$ exposure at this point; instead, it was delayed until after *S* made his $B_2$ estimate (vide infra). For all of the measures analyzed, this variation in treatment produced no significant effects, hence the $A_u$ estimates have been pooled.

## RESULTS

### *The First Stimulus (Card A)*

*Intratrial effects.* In the analysis of these data, two separate measures of influence were utilized. *Movement* is the amount of change from *S*'s initial estimate ($A_1$) to a subsequent estimate for the same stimulus; thus $M_{A_2}$ indicates *S*'s change on Card A from $A_1$ to $A_2$; $M_{A_4}$ indicates the change from $A_1$ to $A_3$ (last estimate). *Conformity* is *S*'s change proportional to the distance between him and his partner; thus, for an *S* who was exposed to a distance of 40 on Card A, $C_{A_2}$ was computed by dividing $M_{A_2}$ by 40; $C_{A_4}$ was computed by dividing $M_{A_4}$ by 40.

Fig. 1 shows the frequency distributions of $M_{A_2}$ for the six distances used (note that the abscissa is scaled geometrically for convenience). At large distances, distributions take on a marked positive skew. Similar skewed distributions obtain for virtually all variables measured after exposure to different distances. Because of this skew, the data were analyzed using the median and other nonparametric statistics.

In Fig. 2, the basic relations between influence and distance are presented. Influence on Photo A is measured by movement and conformity after the first influence attempt in Figs. 2a and 2b, after the last influence attempt in Figs. 2c and 2d. The following facts may be gleaned from the curves:

1. The results are quite stable, as evidenced by the similarity of the data obtained from the two experiments (all medians are based upon an *n* of 32).

2. The median amount of movement increases with distance—at least up to a certain point, after which a decreasing trend is suggested. One might expect, a priori, that for a particular individual, movement will increase with distance until a point of subjective "incredulity" or "implausibility" is reached, above which the influence source will be disre-

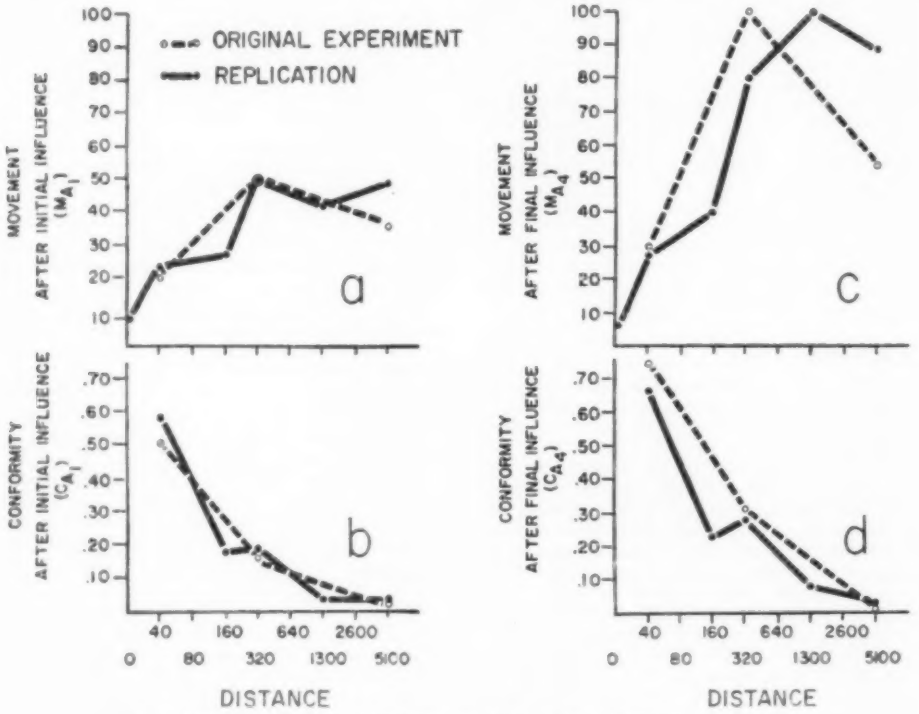


FIG. 2. MEDIAN INFLUENCE MEASURES ON CARD A AS A FUNCTION OF DISTANCE

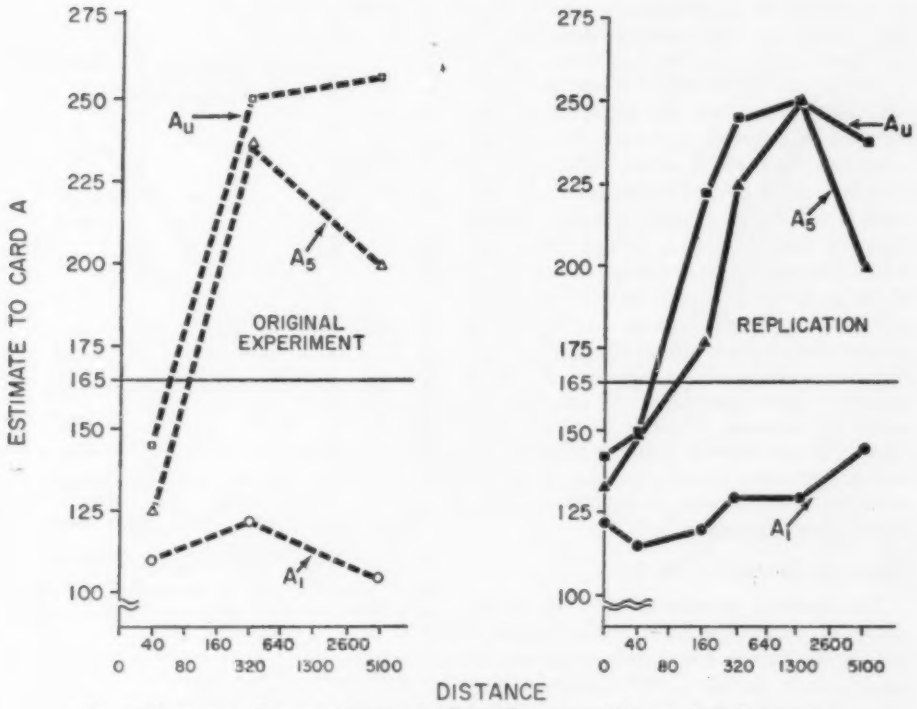


FIG. 3. MEDIANS OF INITIAL ESTIMATE ($A_1$), FIFTH ESTIMATE ($A_5$), AND POST-INFLUENCE ESTIMATE ($A_u$) ON CARD A AS A FUNCTION OF DISTANCE

garded—with movement abruptly dropping to zero.

3. In contrast with movement, conformity is a decreasing nonlinear monotonic function of distance.

4. With repeated influence attempts on the same stimulus, the amount of influence (both movement and conformity) is increased. This is particularly true of movement scores, seen at all distances; $C_{A1}$ and $C_{A4}$, however, are essentially the same at the larger distances. Maximum influence generally occurs on the first influence attempt (cf. 8, 9).

*Persistence of influence.* It will be recalled that after the five estimates on Photo A were "exchanged," *S* was again shown the same photo for 10 seconds and asked to give his "final" estimate ($A_u$). It was expected that, for those *S*s who had been influenced from $A_1$ to $A_5$, an additional exposure to "reality" would permit them to correct themselves and give more accurate estimates. Fig. 3, however, shows that this does not occur. Rather, there is some tendency for the groups *to manifest still further "latent" influence effects on* $A_u$. In both experiments, the median $A_u$ estimates are always equal to or greater than the $A_5$ estimates. Approximately 40 per cent of all *S*s moved from $A_5$ toward the bogus estimate (and *away* from "reality") on $A_u$, while another 10 per cent (whose $A_5$ estimates were all below 165) also moved toward the bogus estimate, coming closer to the true number; 25 per cent of the *S*s did not change their $A_5$ judgments at all, and the remaining 25 per cent moved away from the bogus estimate back toward "reality."

While it appears, then, that influence persists when *S* is re-exposed to the same ambiguous situation (without any direct influence attempt being made), it should not be inferred that these effects represent an enduring change. When *S* is permitted a relatively extensive reality contact with the stimulus (i.e., giving him three minutes to count at the conclusion

of the study), the $A_v$ estimates do tend to drop back toward the true number, regardless of previous distance-experience: *S*s seem capable of shaking off the influence effects to a considerable degree when the ambiguity of the stimulus is removed, although the over-all median $A_v$ (184) still is somewhat higher than the true number (165). This apparent overestimate is difficult to interpret clearly, since an adequate control group is lacking (*S*s in the zero-distance group were exposed to a distance of 320 on Photo B *prior to* their $A_v$ estimates, and therefore do not provide an appropriate comparison); before one could validly infer that the elevation of the obtained $A_v$ curve indicates some persistence of influence, it would be necessary to show that *S*s in the absence of any previous influence attempts are able to estimate accurately (or tend to underestimate) Photo A during a comparable three-minute counting session.

*The Second Stimulus (Card B)*

*The intertrial phenomenon.* In an earlier study of continuous social influence interactions (6), it was found that under certain conditions of immediate self-committal (i.e., where *S* gives an explicit judgment just prior to an influence attempt), a considerable amount of influence may be reflected *between* rather than within interactions (trials). When *S* has been under continuous pressure, applied in some consistent pattern, it is not uncommon to observe *S* appearing to anticipate the influence source on subsequent interactions. Marked changes are seen occurring in *S*'s committal responses themselves, prior to any overt attempt to alter his judgment on any particular stimulus.

It was expected that the intratrial pressures exerted on *S* during the five influence attempts on Photo A would be reflected by a systematic increase in the initial estimate (committal response) to Photo B ($B_1$). From Fig. 4 it is clear that this indeed occurs. $B_1$ is significantly related to the distance *S* was exposed to on Photo A: the larger the distance on Photo A, the larger the median $B_1$. Groups exposed to large distances on a stimulus may be highly resistant to intratrial influence on that particular stimulus, but when exposed to a new stimulus (prior to any direct influence attempt), the previous pressures become patently evident. A number of *S*s in the 5100-distance group showed little intratrial movement on Photo A, yet gave initial estimates of 4500 or higher to Photo B (containing 300 paratroopers).

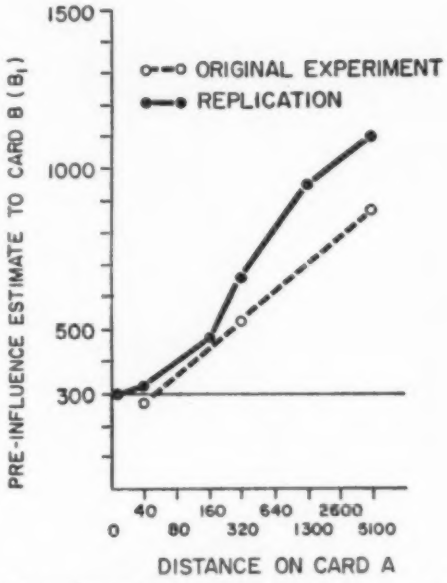


FIG. 4. MEDIAN INITIAL ESTIMATE ($B_1$) ON CARD B AS A FUNCTION OF DISTANCE ON CARD A (INTERTRIAL EFFECT)

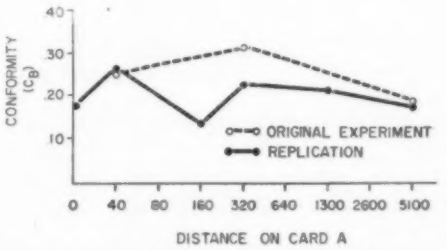


FIG. 5. MEDIAN INTRATRIAL CONFORMITY ($C_B$) ON CARD B AS A FUNCTION OF DISTANCE ON CARD A

*Intratrial effects.* What happens when *S*s who have been exposed to different distances on one stimulus are subsequently exposed to a uniform distance (320) on a second stimulus? It has already been seen that their initial estimate ($B_1$) of this second stimulus is significantly related to their previous distance-experience (intertrial effect). But after they have

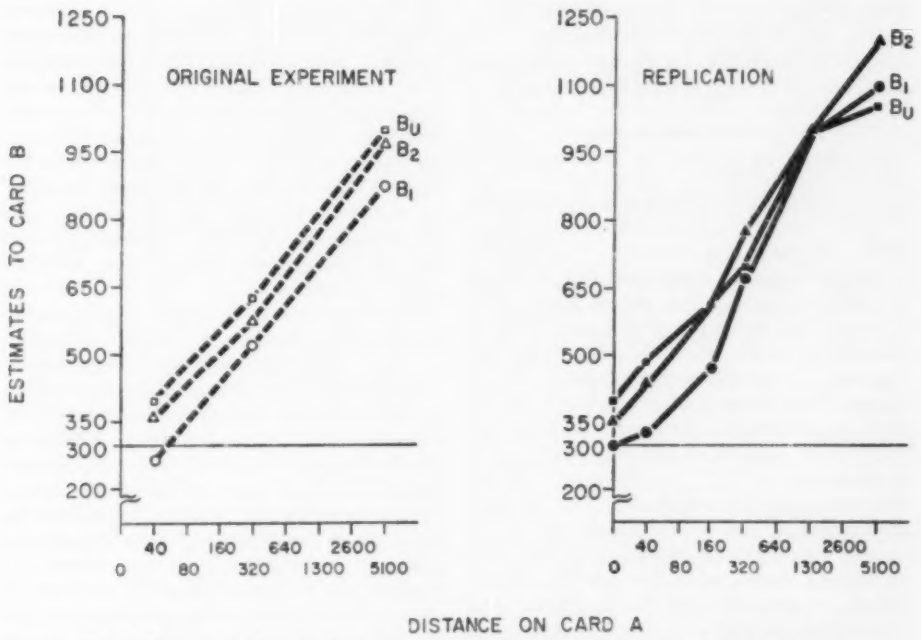


FIG. 6. MEDIANS OF INITIAL ESTIMATE ($B_1$), SECOND ESTIMATE ($B_2$), AND POST-INFLUENCE ESTIMATE ($B_u$) ON CARD B AS A FUNCTION OF DISTANCE ON CARD A

committed themselves to Photo B, how easily can their judgments be further altered? Will *S*s who were exposed to a distance of 5100 on Card A behave differently from *S*s who were exposed to zero distance? Fig. 5 indicates that there is no systematic relationship between $C_B$, the amount of conformity on Photo B, and distance on Photo A. Despite the fact that the 5100-distance group had already demonstrated intertrial influence on Photo B, this group appears to change neither more nor less than any other group—including the zero-distance *S*s who show essentially no intertrial effects.

The rank-order correlations of $C_B$ with several other variables (e.g., $C_{A1}$, $A_1$, $A_5$, $B_1$) were computed within each distance-group, but no significant relations were obtained. Under the conditions of the present study, then, $C_B$ cannot be demonstrated to be related to any of the temporally previous variables measured, including distance on Photo A.

*Persistence of influence.* When *S*s were given a ten-second re-exposure to Photo B ($B_u$), Fig. 6 reveals a marked tendency toward maintenance of the previous influence effects, similar to the results on Photo A. The $B_u$ curve closely follows the $B_2$ curve at all distances: *S*s do *not* show any strong indication of dropping back to "reality" even at the extreme distances. However, the "latent" influence effect on $B_u$ is not as clear as for $A_u$.

When given three minutes to count Photo B at the conclusion of the study, *S*s' estimates ($B_v$) are now independent of the distance on Photo A. All groups have similar median $B_v$ judgments (approximately 500), but all medians are considerably above the true number of paratroopers (300). As noted in the case of $A_v$, interpretation is equivocal in the absence of proper controls, but there is certainly the suggestion that the "counting" behavior of these *S*s may be modified by previously applied pressures. Since it was virtually impossible to make an accurate count of the 300 paratroopers on Photo B (quite different from Photo A, which contained only about half the number), the $B_v$ measurement becomes in effect a continued influence situation—differing from $B_u$ (seen for only 10 seconds) only in the degree of stimulus ambiguity. A few *S*s in the

1300 and 5100 distance groups who had given high $B_1$ and $B_u$ estimates were observed to go through the complete motions of "counting," and then give $B_v$ judgments of 3500 or 4000!

## DISCUSSION

### *The Effects of Distance*

One of the major questions in this research concerned the relationship between measures of social influence and distance. The results clearly indicate that while movement tends to increase (at least up to a certain point) with distance, conformity is a monotonic nonlinear decreasing function. It would seem, then, that distance is a significant determinant of social influence, and hence must be considered in any attempt to formulate theoretical models of influence—particularly since a very recent paper by Hovland and Pritzker (11) reports similar relations using attitudinal stimuli low in ego-involvement. These results do not appear to support Goldberg's earlier finding (8) that conformity is independent of distance. The present results are also pertinent to the Osgood and Tannenbaum model (19), French's theory (7), and Mausner's (17) recent discussion.

### *Intertrial Effects and Persistence of Influence*

A striking phenomenon consistently observed in recent studies of social influence in continuous interactions (6) concerns the large amounts of influence manifested *between* interactions (i.e., in this instance, stimuli A and B). In the present experiment, *S*s on Photo A not only tended to remain influenced when the *same* stimulus was presented a second time ($A_u$), but also showed a carry-over effect when a new stimulus (Photo B) was presented: *S*s demonstrated gross over-estimation tendencies positively related to the distance-pressure received on the first stimulus. These intertrial effects, occurring prior to any overt, direct influence attempts on the second stimulus, often seem to rest in "anticipatory" attitudes in *S*. Just as expectations play an important role in hypnotic behavior (3, 4), similar factors seem to be operative in nonhypnotic continuous influence situations. To the extent that *S* can perceive a more or less consistent bias in the source's judgments, his frame of reference (22) is altered, and he subsequently tends to manifest indirect influence effects in giving his initial, pre-influence (committal) judgments to new stimuli. Some *S*s may even be quite resistant to intratrial pressures on the first stimulus, and then show a large intertrial effect. (This phenomenon may be a generalization of the "sleeper effect," noted in a number of attitude-influence studies [e.g., 10, 13]: the main difference seems to be that whereas the "sleeper effect" refers to increased influence observed in subsequent retesting on the *same* stimulus, the intertrial effect refers to increased influence seen during testing on a *new* but related stimulus.)

When the stimuli all belong to a quantitative dimension, as in the present study, it is probable that some psychophysical "anchoring" effects from *S*'s previous judgments partly contribute to the intertrial change—i.e., if an *S* was actually influenced on Photo A, then he will overestimate Photo B merely to remain consistent within his newly established frame of reference; this sort of behavior need not be described as "anticipatory," since it is similar to the type of judgment change produced by Tresselt and Volkman (24) in the absence of social interaction. However, as Sherif (19, 20) has pointed out, both anchoring and anticipatory judgments have in common a shift in *S*'s frame of reference. Thus, in some cases, *S* may be anticipating his partner, attempting to "beat him to the punch" (i.e., minimize the anticipated distance between himself and the influence source). In other cases, *S* may still be "anticipating" the experimental situation, which requires a new frame of reference based upon *S*'s previous experience in the situation.

### *Subsequent Intratrial Influence*

The finding that conformity on Photo B is independent of the distance on Photo A raises a question. The present results seem to indicate that distance on the first interaction is a basic determinant only of (*a*) intratrial movement on that first interaction, and (*b*) intertrial movement between the first and second interactions. Can it be concluded, then, that in real life continuous social influence situations, the distance-exposure of an *S* on one interaction has no predictive power in regard to his intratrial behavior in a subsequent interaction on a related issue? This is quite unlikely, when one considers the results of a variety of experiments on "prior reinforcement."

When *S*s, following an initial interaction (or

series), have an opportunity to *learn* something about the correctness of their judgments and/or those of their partners (i.e., the influence source), this learning then affects their subsequent judgments differentially (12, 14, 15, 16, 17, 21). For example, if there is some sort of feedback (knowledge of results, or "validation") so that *S* learns he was "right" and/or partner "wrong" on the initial interaction, then subsequent intratrial conformity will be modified. Goldberg and Lubin (9) further found that, when *S* is given some validation information, the distance on the preliminary interactions becomes a significant determinant of subsequent intratrial conformity.

Yet the data from the present study show no such effects: *S*s interacting with a peer partner at different distances on one interaction do not show differential conformity on the next interaction. These findings suggest that, unless *S* can "validate" his own and/or the influence source's judgments in any initial interaction, his intratrial conformity behavior on a second interaction will be independent of the distance he received on the first interaction. Thus, the present study indicates that the variable of distance plays a salient role not only on a particular influence attempt, but in *continuous* social influence processes. (It has also been suggested [5] that distance may be a fruitful concept in the analysis of related influence processes such as psychotherapy.)

## SUMMARY AND CONCLUSIONS

Pairs of *S*s were placed in a social influence situation requiring judgments of the number of paratroopers seen in briefly exposed photographs. Two photos were used successively (Cards A and B) after a preliminary warm-up period. On Photo A, distance (the size of the discrepancy between *S*'s judgment and his partner's) was systematically varied by intercepting *S*'s written estimate and replacing it with a bogus estimate at a predetermined distance. After a number of judgments were obtained on Card A, the *S*s were exposed to Card B. On this latter photo, all *S*s were exposed to the same distance. The experiment was replicated with the following major results:

1. At large distances, distributions of the amount of influence on Photo A take on marked positive skewness, the amount of skew increasing with distance.

2. The amount of influence on Photo A is significantly related to distance: (*a*) Movement (the amount of change from initial judgment) increases with distance up to a certain point; there is a slight suggestion (supported by rational consideration) that the curve begins to decrease at extremely large distances. (*b*) Conformity (movement relative to distance) decreases with distance in a monotonic fashion.

3. The amount of influence obtained on Photo A can be increased by sheer repetition of the suggestion, in the absence of any "side arguments," persuasive reasons, or elaborate rationales.

4. When *S*s are permitted a second brief exposure to the stimuli, there is no evidence that *S*s "correct" their previous post-influence judgments toward "reality." In fact, on Card A in particular, the data indicate that *S*s tend to show increased "latent" influence on the re-exposure. However, when *S*s are permitted to count the paratroopers at the end of the experiment, the judgments are no longer related to distance. Part of the earlier influence effects may carry over to this "counting" behavior, depending upon the ambiguity of the stimuli.

5. Considerable amounts of influence are manifested *between* interactions on successive stimuli (intertrial effect). Marked changes are seen in *S*'s initial response to the second stimulus, before any overt attempt to alter his judgment on this particular photo. The intertrial effects are positively related to the distance on the first stimulus. A number of *S*s who were resistant to intratrial influence at large distances on Photo A showed considerable intertrial change.

6. Intratrial influence on the second stimulus was found to be independent of the distance on the first stimulus. It was hypothesized that unless *S* can "validate" his own and/or his partner's judgments in an initial interaction, his intratrial influence behavior on a second interaction will be unrelated to the distance on the first interaction.

## REFERENCES

1. ASCH, S. E. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**, (Whole No. 416).

2. FESTINGER, L., GERARD, H. B., HYMOVITCH, B., KELLEY, H. H., & RAVEN, B. The influence

process in the presence of extreme deviates. *Hum. Relat.*, 1952, **5**, 327–346.

3. FISHER, S. The role of expectancy in the performance of posthypnotic behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 503–507.
4. FISHER, S. An investigation of alleged conditioning phenomena under hypnosis. *J. clin. exper. Hypnosis*, 1955, **3**, 71–103.
5. FISHER, S. Plausibility and depth of interpretation. *J. consult. Psychol.*, 1956, **20**, 249–256.
6. FISHER, S., RUBINSTEIN, I., & FREEMAN, R. W. Intertrial effects of immediate self-committal in a continuous social influence situation. *J. abnorm. soc. Psychol.*, 1956, **52**, 200–207.
7. FRENCH, J. R. P. A formal theory of social power. *Psychol. Rev.*, 1956, **63**, 181–194.
8. GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.
9. GOLDBERG, S. C., & LUBIN, A. Influence as a function of prior judgment error. Paper read at American Psychol. Ass., New York City, September, 1957.
10. HOVLAND, C. I., & WEISS, W. The influence of source credibility on communication effectiveness. *Publ. Opin. Quart.*, 1952, **15**, 635–650.
11. HOVLAND, C. I., & PRITZKER, H. A. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psychol.*, 1957, **54**, 257–261.
12. KELMAN, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267–285.
13. KELMAN, H. C., & HOVLAND, C. I. "Reinstatement" of the communicator in delayed measurement of opinion change. *J. abnorm. soc. Psychol.*, 1953, **48**, 327–335.
14. LUCHINS, A. S., & LUCHINS, E. H. Previous experience with ambiguous and nonambiguous perceptual stimuli under various social influences. *J. soc. Psychol.*, 1955, **42**, 249–270.
15. MAUSNER, B. The effect of prior reinforcement on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 65–68.
16. MAUSNER, B. The effect of one partner's success in a relevant task on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 557–560.
17. MAUSNER, B. Studies in social interaction: I. A conceptual scheme. *J. soc. Psychology*, 1955, **41**, 259–270.
18. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.
19. OSGOOD, C. E. &. TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.
20. SCHACHTER, S. Deviation, rejection and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.
21. SCHEIN, E. H. The effect of reward on adult imitative behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 389–395.
22. SHERIF, M. *The psychology of social norms*. New York: Harper, 1936.
23. SHERIF, M., & CANTRIL, H. The psychology of attitudes. *Psychol. Rev.*, 1945, **22**, 295–319.
24. TRESSELT, MARGARET E., & VOLKMANN, J. The production of uniform opinion by nonsocial stimulation. *J. abnorm. soc. Psychol.*, 1942, **37**, 234–243.

Received April 24, 1957.

# THE INFLUENCE OF LANGUAGE-ACQUISITION CONTEXTS ON BILINGUALISM[1]

W. E. LAMBERT, J. HAVELKA,[2] AND C. CROSBY

*McGill University*

THE bilingual person presumably must learn two different verbal responses for each referent, a requirement analogous to learning different responses to the same stimulus in experimental studies of negative transfer. It is of psychological interest, therefore, that bilinguals are so often able to communicate freely in either language without interference from the other. As yet, no adequate psychological theory has been offered to account for bilingualism. In his review of various limited explanations, Weinreich concludes that "a comprehensive psychological theory of bilingualism ought to account for both the effectively separated use of the two languages and for interference of the languages with one another" (9, p. 71). It is the purpose of the present investigation to contribute to such a theory by analyzing conditions that are presumed to affect both the "effectively separated use" of the bilingual's two languages and the "interference of the languages with one another."

As a first step, it may be profitable to apply certain principles derived from paired-associate learning and transfer phenomena to the case of learning two languages.[3] In the experimental study of retroactive inhibition, the learner must first link $S_1$ to $R_1$ and, in the later interpolated phase, $S_1$ to $R_2$. When alternative responses are associated with a common stimulus, interference between the two associations is the usual consequence. The bilingual is confronted with an analogous learning problem when, for example, the symbols "church" and "*église*" are both associated with the appropriate environmental event.

One general means of minimizing interference in learning situations is to increase the distinctiveness of the two tasks to be learned.[4] In studies of retroactive inhibition, for example, interference is effectively minimized by increasing the degree of learning of original and interpolated tasks or by decreasing the associative similarity between the competing responses. Both these variations enhance the distinctiveness of the tasks. If distinctive secondary cues, such as differently colored backgrounds, are associated with the two tasks to be learned, the functional separation of the tasks should be increased.

Following the logic of interference theory, we can reason that response competition in bilingualism is functionally related to the distinctiveness of contexts in which the bilingual's two languages are acquired; the more "separated" the contexts of acquisition, the less the bilingual interference. Acquisition contexts are considered to be separated when the association between a symbol such as "church" and the appropriate environmental event has consistently taken place in a setting distinct from that in which the association between "*église*" and its corresponding environmental event occurs. On the other hand, in "fused" acquisition contexts, members of a linguistic community can and do use both symbols for any particular environmental event, thereby increasing the interchangeability of the symbols.

Recently, Ervin and Osgood (4) attempted a theoretical interpretation of bilingualism that indicates the potential importance of language-acquisition contexts. They suggest that bilinguals could theoretically acquire "compound" or "coordinate" language systems. The compound system would be developed through ex-

[1] This research was supported in part by the Canadian Defence Research Board, Grant Number D77-94-01-10, and in part by N.R.C. Grant A.P. 17.

[2] Now at the University of Western Ontario.

[3] In 1915, Epstein (2) considered bilingualism in terms of such a psychological theory, but did not experimentally test its applicability.

[4] Essentially the same reasoning underlies the "direct method" of language teaching. The direct method theorists maintain that the acquired language should be learned directly from association with environmental events without the mediation of the native language. When learning French as a second language, one should never link "pomme" with "apple" but should link "pomme" directly with the perceived environmental event. In contrast to the "indirect" method which utilizes translation as an aid in acquisition, the direct method is purported to develop "pure" bilingualism, the ability to move from one linguistic and semantic world of experience to another with comparatively little interference.

perience in fused contexts, as with vocabulary training in school, or where the same family members use two languages interchangeably to refer to the same environmental events. The coordinate system would be developed through experience in different linguistic communities where languages are rarely interchanged. They suggest that the alternative symbols in his two languages have a single meaning for the compound bilingual whereas the coordinate bilingual, presumably because of the distinctiveness of his language-acquisition contexts, develops separate meanings for each of his alternative symbols.

Accepting this theoretical position, we hypothesized that experience in "separated" language-acquisition contexts enhances the functional separation of the bilingual's two languages while experience in "fused" contexts reduces the functional separation of the two language systems. We predicted that bilinguals with experience in separated contexts would show comparatively greater semantic differences between symbols in one language and their translated equivalents in the other, more associative independence of translated equivalents in their two languages, and less facility of switching from one language to the other.

## METHOD

### *Classification of Subjects*

The *S*s were college or university students with extensive experience in both English and French languages. Since all but four were either preparing for or actually in graduate schools, it can be assumed that the group was relatively homogeneous in intelligence. Each *S* was tested for extent of bilingualism (5, 6) and the 32 who met our criterion for bilingual "balance" will be discussed here. "Balance" implies a high level of second language competence, high enough, in fact, that no dominance of one language over the other is measurable with the technique employed.

Our primary task was to classify *S*s according to their language-acquisition contexts and, to this end, detailed information was obtained from each *S* concerning how, when, and where his languages were acquired. Acquisition contexts were considered *separated* when one language was learned exclusively in the home and the other exclusively outside the home, when one parent consistently used one language and the other parent a different language, or when one language was acquired in a particular national or cultural setting distinct from that in which the second was acquired. Acquisition contexts were considered *fused* when both parents used the two languages indiscriminately, when both languages were used interchangeably inside and outside the home, or when an individual acquired his second language in a school system stressing vocabulary drill and translation and where the first or native language was used as the medium of instruction. This scheme of classification is derived from that used by Ervin and Osgood (4) in their definitions of compound and coordinate systems. The actual classification of *S*s as having separated or fused contexts for acquisition is essentially an arbitrary one in spite of the rather explicit criteria for placement. In five cases, for example, a problem arose in deciding whether to give credit for some experience with separated contexts when an *S* had early experience in a fused context, as when an English *S* acquired French in a school system but later used this language exclusively in France for some period of time. Our decision was to classify such an *S* as separated if he spent at least an entire year using the acquired language either exclusively or primarily. Using this scheme, 23 *S*s were classified as "separated" and 9 as "fused."[5]

It was also possible to further subdivide the "separated" group according to the following line of reasoning. An *S* could be classified as "separated" if he had experience either in distinctive acquisition contexts within a particular national setting, as in Quebec, or in two geographically different cultures, as from living many years in France before coming to America. We predicted that separated contexts based on "bicultural" experiences, as in the latter example, would particularly enhance the distinctiveness of the two contexts. In this case, the environmental events themselves may often not be common for the two sets of symbols. For example, "*église*" could mean a gothic cathedral while "church" could mean a tall

[5] In view of our method of selecting *S*s (see 6) where an attempt was made to draw equal numbers of both types, one may safely generalize that it is comparatively easier for bilinguals with separated acquisition contexts to attain bilingual balance.

wooden building used on Sundays, and *"ami"* and "friend" could take on quite different meanings. A recent study by Ervin (3) demonstrates how such environmental differences affect the perceptual and personality functionings of bilinguals. With these considerations in mind, the "separated" group was subdivided into 15 *S*s with bicultural and 8 *S*s with "unicultural" experiences.

*Behavioral Indices of the Influence of Acquisition Contexts*

As a measure of the differences in meanings of concepts in the bilingual's two languages, we made use of Osgood's "semantic differential" (7), a method of scaling stimulus-words on a standard set of meaning dimensions. For example, an *S* is presented with a stimulus-word, such as "house," and is asked to rate it along a seven-point dimension, such as "fast-slow" or "pleasant-unpleasant." In the present case, the stimulus-words were four common English concepts ("house," "drink," "poor," "me") and their French translations. They were presented in a fixed order so that any word was maximally separated from its translated pair. Ten standard dimensions (see 8) were used and their order of presentation was continually changed to eliminate any patterning of ratings. The *S* might rate "house" at Position 3 on the "fast-slow" dimension, and later "maison" might be rated at Position 5 (more slow) on the same dimension, which had been translated into French. Since both stimulus-words and the anchors of each dimension were translated, we can only make statements about differences in meanings of both stimulus-words and dimensional attributes, or their interaction. The average degree of difference between the French and English semantic ratings for the four pairs of stimulus words was determined for each *S* using Osgood's *D* score (7); the larger the *D* score, the greater the semantic differences between translated equivalents in French and English. We predicted that *S*s with experience in separated language-acquisition contexts would show comparatively greater semantic differences between translated equivalents in their two languages.

As a measure of the degree of associative independence of translated equivalents, we made use of the retroactive inhibition design. If an *S* learns list A, then list B, then re-learns list A, one can determine the amount of associative interference imposed by the interpolated material. If the interpolation has no effect on the previously memorized material, then the two can be considered functionally independent. Interference and facilitation from the interpolated material may be thought of as indices of the functional dependence of the first-learned and interpolated tasks.

Each *S* learned a list of 20 commonly occurring English words by the anticipation method. A criterion of 8 correct responses out of 20 was chosen so as to permit improvement following the interpolated task. This was followed by three presentations of a list of 20 three-letter nonsense words. Then the English list was re-presented and the *S*'s retention was measured. Finally, *S*s were asked to relearn the nonsense list, but these data will not be discussed here. This preliminary test gives a measure of the associative interference or facilitation imposed by a meaningless task, one assumed to be semantically independent of the original task. After a rest period, the same procedure was repeated, but this time a second, similar list of 20 English words, learned to the same criterion, was followed by a list of their exact translations in French, presented for three trials. A score was assigned to each *S* which indicated the degree of associative interference or facilitation imposed by the interpolated French material in relation to interpolated meaningless material. The following illustrative case indicates some facilitation from both interpolated tasks but more from the interpolated French. The number of correct responses for the criterion trials are placed in Column A, and the number of correct responses for retention trials are given in Column A′, for the English-Nonsense-English and the English-French-English series:

| | A | I | A′ | $A - A'$ |
|---|---|---|---|---|
| Series I: | $E_1 = 10$ | *N*(3 trials) | $E_1 = 11$ | −1 |
| Series II: | $E_2 = 10$ | *F*(3 trials) | $E_2 = 15$ | −5 |
| | | I–II = +4 | | |

Positive scores indicate that the interpolated French translations facilitate retention of English words more than do interpolated nonsense words. Negative scores ind cate that interpolated French translations interfere with the

retention of English words. A zero score would indicate that the interpolated French translations offer no more facilitation than interpolated nonsense words. Following the reasoning outlined above, the closer the scores approach zero, the more associatively independent the translated equivalents. We predicted that *S*s with experience in separated acquisition contexts would show more associative independence between translated equivalents in their two languages than those with experience in fused contexts.

Finally, we predicted that experience in fused as compared to separated contexts would express itself in a greater facility of switching from one language to the other. As a measure of such facility, we made use of a speed of translation test described elsewhere (6). Each *S*, given 20 French and 20 English words to translate, was assigned a score for his average speed of translation from English to French and from French to English. Speed was measured by a voice key in circuit with a chronoscope.

## RESULTS AND DISCUSSION

As predicted, those bilinguals who acquired their languages in separated contexts showed a significantly greater difference in meanings of translated equivalents than did those who acquired their two languages in fused contexts (see Table 1). However, when a comparison was made between those who acquired their two languages in geographically distinct cultures (bicultural experience) and those who acquired both languages in separated contexts within one geographical region (unicultural experience), it became clear that bicultural experience accounted for the over-all difference between fused and separated groups. In fact, *S*s with unicultural experience are not essentially different from either of the other two groups with respect to similarity of meanings of translated equivalents. It appears, therefore, that the difference in meanings of translated equivalents is not a simple function of experience in separated acquisition contexts. Experience in bicultural acquisition contexts, where there is a greater likelihood that the actual referents of translated symbols are different, does appear to affect the semantic aspects of the bilingual's two languages.

Bilinguals with experience in separated acquisition contexts also showed significantly more associative independence of translated equivalents in their two languages than did those with experience in fused contexts (see Table 2). An interpolated list of French translations had about the same influence for the separated group as an interpolated list of nonsense syllables. *S*s in the fused group, on the other hand, clearly benefited from the interpolated French list.

These findings are consistent with those summarized above for semantic differences. Since translated equivalents were semantically more similar for the fused *S*s, the interpolated translations in the learning problem were more like an extension of the original English list. It is logical therefore that these *S*s were helped more by the interpolated French than were the separated *S*s. Since translated equivalents have comparatively different meanings for the separated *S*s, interpolated translations should have shown less associative facilitation and possibly more interference with the original English list, as was the case.

This interpretation has its limitations, how-

TABLE 1

SEMANTIC DIFFERENTIAL COMPARISONS

| Contexts | *N* | Mean English-French Profile Separations[a] | Variance | | *t* tests[b] | *p* values |
|---|---|---|---|---|---|---|
| Fused (F) | 9 | 3.19 | 0.38 | F-S: | 2.24 | <.05 |
| Separated (S) | 23 | 3.93 | 1.23 | F-B: | 2.64 | <.03 |
| Bicultural (B) | 15 | 4.22 | 1.47 | F-U: | <1 | n.s. |
| Unicultural (U) | 8 | 3.39 | 0.33 | B-U: | 2.13 | <.06 |

[a] Entries are mean *D* scores (7). The larger the mean, the greater the semantic differences between languages.

[b] Where the variances were not homogeneous, *p* values were determined according to a formula suggested by Edwards (1, p. 273 ff). All *t* tests in this and the following tables are 2-tailed tests of significance.

TABLE 2

ASSOCIATIVE-INDEPENDENCE COMPARISONS

| Contexts | *N* | Mean Independence Indices[a] | Variance | | *t* tests | *p* values |
|---|---|---|---|---|---|---|
| Fused (F) | 9 | +3.44 | 9.75 | F-S: | 2.63 | <.02 |
| Separated (S) | 23 | +0.13 | 11.23 | F-B: | 2.12 | <.05 |
| Bicultural (B) | 15 | +0.60 | 10.86 | F-U: | 2.55 | <.02 |
| Unicultural (U) | 8 | −0.75 | 12.28 | B-U: | <1 | n.s. |

[a] The closer the scores approach zero, the more between-language associative independence indicated. A constant of 8 was added to each score to eliminate negative values before variances were computed.

ever, since the data in Table 2 indicate that there is no difference in degree of associative independence between bilinguals with bicultural and those with unicultural experiences. In fact, both these subgroup means are significantly different from that of the fused group. This finding suggests that experience in separated contexts of either a bicultural or unicultural nature enhances the associative independence of translated equivalents in the bilingual's two languages, as compared to experience in fused acquisition contexts.

Table 3 makes it clear that there were no differences between groups in facility of switching from one language to the other, a finding contrary to predictions. There is no immediately obvious interpretation for this lack of difference between groups, other than the possible lack of sensitivity of our measuring device. Accepting the finding as reliable, one may speculate that *S*s in both groups have had equal experience in translating but possibly different types of experience. The fused *S*s, by definition, have learned their two languages in contexts that encourage translation. On the other hand, a bilingual who has contacts with two separate linguistic communities may often be called on to recount his experiences with members of both communities. In a sense, he becomes the liaison between two linguistic communities and he may learn to use concepts which have meaningful equivalents for members of both communities.

When requested to translate, *S*s in the two groups demonstrate similar facility to switch. In view of the other findings, however, the fused and separated *S*s appear to make different use of the translation process. The fused *S*s made use of the interpolated translations in the learning task and accordingly showed improvement in their retention scores. It is possible, although no attention was paid to this feature in the collection of data, that the separated *S*s were less aware that actual translations were being interpolated than were the fused *S*s. Stated differently, the fused *S*s may have been more set to translate than were the separated *S*s.

In conclusion, we have been able to demonstrate, within the limitations of the small sample of *S*s studied, that language-acquisition context does influence the functional separation of the bilingual's two languages in certain respects. The theory of coordinate and compound language systems has been given empirical support and the defining characteristics of these systems have been extended. The coordinate bilinguals, in contrast to the compound bilinguals, appear to have more functionally independent language systems. If their coordinateness has been developed through experience in culturally distinctive contexts, they will have comparatively different meanings for concepts translated into their two languages. Coordinate and compound bilinguals, however, appear to have equal facility in switching from one language to the other.

TABLE 3
SWITCHING FACILITY COMPARISONS

| Contexts | *N* | Mean Translation Speed[a] | Variance | | *t* test | *p* values |
|---|---|---|---|---|---|---|
| Fused (F)[b] | 7 | 125.08 | 112.43 | F-S: | <1 | n.s. |
| Separated (S) | 23 | 125.69 | 288.65 | F-B: | <1 | n.s. |
| Bicultural (B) | 15 | 124.59 | 303.42 | F-U: | <1 | n.s. |
| Unicultural (U) | 8 | 127.76 | 254.62 | B-U: | <1 | n.s. |

[a] Entries, in hundredths of a second, are group means. The larger the mean, the slower the group is in translating into French and into English.

[b] Two *S*s were not available for this test.

## SUMMARY

Bilinguals were classified as having learned their two languages in either separated or fused contexts. It was hypothesized that experience in separated as compared with fused language-acquisition contexts comparatively enhances the effectively separated use of the bilingual's two languages.

It was found that experience in separated contexts comparatively increases the associative independence of translated equivalents in the bilingual's two languages. If the bilingual has learned his two languages in culturally distinctive contexts, the semantic differences between translated equivalents is comparatively increased. There was no difference found in facility to switch from one language to the other that can be attributed to contextual influences.

The findings were related to the theory of compound and coordinate bilingual language systems.

## REFERENCES

1. EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.

2. EPSTEIN, I. *La pensée et la polyglossie.* Paris: Payot, approx. 1915.
3. ERVIN, S. M. The verbal behavior of bilinguals: the effect of the language of report upon the T.A.T. stories of adult French bilinguals. Unpublished dissertation, Univer. of Michigan, 1955.
4. ERVIN, S. M., & OSGOOD, C. E. Second language learning and bilingualism. In C. E. Osgood and F. Sebeok (Eds.), *Psycholinguistics. J. abnorm. soc. Psychol., Suppl.*, 1954. Pp. 139–146.
5. LAMBERT, W. E. Measurement of the linguistic dominance of bilinguals. *J. abnorm. soc. Psychol.*, 1955, **50**, 197–200.
6. LAMBERT, W. E., HAVELKA, J., & GARDNER, R. Linguistic manifestations of bilingualism. (Mimeo.)
7. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.
8. OSGOOD, C. E., & LURIA, Z. A blind analysis of a case of multiple personality using the semantic differential. *J. abnorm. soc. Psychol.*, 1954, **49**, 579–591.
9. WEINREICH, U. *Languages in contact.* New York: Linguistic Circle of New York, 1953.

Received April 23, 1957.

# OBLIGATIONS TO SELF AND SOCIETY IN THE UNITED STATES AND GERMANY[1]

DAVID C. McCLELLAND, J. F. STURR,[2] R. H. KNAPP, AND H. W. WENDT[3]

*Harvard University and Wesleyan University*

MUCH has been written about the conscious ideals and unconscious motives of both Germans and Americans (3, 4, 5, 7, 8, 9, 11, 13, 14, 16), but little of it has been based on factual evidence systematically gathered for comparative purposes. The present study does not pretend to be complete or definitive because it is based on a very limited sample of upper middle-class male adolescents, but it does attempt (*a*) to see the value systems of the United States and Germany as functioning wholes and to compare them without prejudice as to which is better, and (*b*) to make the comparisons on the basis of actual measurements taken from individual subjects.

While it is fair to state that it was the intention to make the comparison without prejudice, it is also true that no study can be made without preconceptions. So it would be well to state these in advance as explicitly as possible: the general assumption which guided the study was that every society has to strike some kind of a balance between freedom and order, between the wishes of the individual and the obligations of social living, between egocentricism and sociocentricism. Such an assumption has particular value for this study for two reasons. First, two of the best-developed measures of human motivation—n Achievement and n Affiliation—fall easily into one orientation or the other: n Achievement into the egocentric and n Affiliation into the sociocentric orientation. And, second, previous research as well as popular stereotype has tended to picture Americans as "individualistic" (with over-tones of "other-direction" and conformity, to be sure) and the Germans as "sociocentric" (again with a highly individualized sense of honor). So, while this study was designed to be an exploratory inductive enterprise, it was guided by the very general notion that measures should be used that would contrast the two cultures on the obligations to self and to society, although, as the ensuing section will show, many measures were also taken that did not fit precisely into this scheme.

[1] This study was completed when all the authors were connected with Wesleyan University and is based on J. F. Sturr: A comparison of the values of German and American secondary school leaders and non-leaders. Unpublished A.B. thesis, Honors College, Wesleyan University, Middletown, Connecticut, 1955.

[2] Now at the University of Rochester.

[3] Now at the University of Mainz.

## METHOD

### *Measures*

The materials used consisted of a picture-story test of n Affiliation and n Achievement and a questionnaire. The method of deriving n Achievement and n Affiliation scores from brief written stories to pictures has been fully described elsewhere (12, 17). In this study, the pictures used to elicit stories related to achievement were: a male adolescent at a desk, a foreman and worker in a factory, a "father" talking to his "son" with a farm background, and a man working at a desk in his office. The pictures used for measuring n Affiliation were: seven men seated around a table, a mother-son scene (TAT 6BM), a man standing alone looking out through a door, and a figure underneath a lamp post at night (TAT 20). The instructions for eliciting the stories were standard (12) and so far as possible identical in the two countries. The coding of the German stories was done in German using a translation of the English manuals for scoring n Achievement and n Affiliation.

The questionnaire consisted of 38 items of the F-scale type (1) to which the subject (*S*) was to express the degree of his agreement or disagreement on a scale from +3 (for complete agreement) to −3 (for complete disagreement). After the questionnaires were returned, the scale was changed to run from +1 for −3 to 7 for +3, as is common practice, all the results being reported in the latter terms. Three basic orientations influenced the choice of items for inclusion in the questionnaire—the self vs. society dimension already mentioned, previous speculative or research findings on differences between Germans and Americans, and the framework set up by F. Kluckhohn (10) for the comparative study of values. The variety of sentiments covered is illustrated by the fact that at least two items touched on each of the following ideas: achieved vs. ascribed status, loyalty to family and country, emotional control or pride, self-development, kindness and love, community service, affiliative values (friendship), achievement values, other-directedness, man's relationship to nature (e.g., as "over" or "under" nature), the importance of fate and destiny, orderliness and rational planning, and will power. In some cases, a single item was believed to tap more than one of these ideas so that the categories were in no sense conceived

as mutually exclusive. In fact, since the categories will play very little part in our analysis, the chief point in mentioning them is to demonstrate how wide a net was cast and to make it easier for subsequent investigators to find obvious areas for comparison that were overlooked.

There were also three open-ended questions at the end of the schedule. The first asked the student to list his extracurricular activities (clubs, hobbies, sports, etc.), which were then simply counted and coded as involving either group or individual activities. The former included all clubs and team sports (soccer, football, basketball), and the latter individual hobbies (e.g., stamp-collecting) and individual sports (hiking, playing chess or tennis). The second open-ended question asked "What are the three things you would most like to teach your children?" and the answers were coded as being either egocentric (e.g., to do well in school, to appreciate music, to enjoy life) or sociocentric (e.g., to be kind, respect others, to be loyal to one's friends). The final open-ended question asked them to list the "six students in your class who you think are most likely to be leading members of their community 30 years from now." The purpose of this item was to break the students roughly into two groups—one consisting of "leaders" and the other of "nonleaders" at least from the point of view of their peers. In all classes tested, both in Germany and the United States, it turned out that roughly half the *Ss* received three or more votes and so this cutting point was used to divide the *Ss* into "leaders" and "nonleaders," although it should be clear that the categorization is very crude, and that the term "leadership" is being used in a very restricted sense in this context.

### *Subjects*

The *Ss* used were males nearly all between the ages of 16 and 19 (mode 17, median nearly 18 in both countries). Furthermore, all of them belonged to what might be called the educational élite who were preparing to go on for higher education either in a college or university. The nature of the sample was dictated to some extent by the way the German educational system is organized. Beginning at the age of 10, German students who are going on to the university tend to be segregated from students with lesser educational or occupational goals (who go to *Volksschule* or *Mittelschule*), so that a sample of German students tested in the "higher" schools around the age of 18 tend to be a more highly selected group than one would find in the ordinary United States high school senior class. Since it was difficult to get a cross-section of German students at this age, it was decided to limit the comparison to the relatively élite group to be found in German higher schools. Fifty-one students were obtained from a Neusprachliches Gymnasium and 36 from an Altsprachliches Gymnasium, in Kaiserslautern, Germany. To match this sample, *Ss* were drawn from college preparatory classes in a high school in Utica, New York ($N = 35$) and from a New England private school ($N = 39$). The private school students were tested to provide a group more nearly comparable to those in the Altsprachliches Gymnasium which emphasized a classical curriculum, and also to make certain that the sample of United States *Ss* was as highly selected as the German one. They also make the American sample unrepresentative because private schools exist in large numbers only in the northeastern section of the United States, but the purpose of the study is not primarily to get representativeness of each culture but to contrast the value orientations of comparable young male élites within the two cultures. The testing was done in December, 1954 in the United States and in March, 1955 in Germany.

TABLE 1

DISTRIBUTION OF OCCUPATIONAL LEVELS OF FATHERS OF BOYS IN UNITED STATES AND GERMAN SAMPLES

| Occupational Level | U.S. ($N$ = 70)[a] | Germany ($N$ = 85)[a] |
|---|---|---|
| 6. Major professional and business executive | 32% | 24% |
| 5. Minor professional and medium business | 26 | 39 |
| 6 and 5 combined | 58% | 63% |
| 4. Clerical and sales | 24 | 22 |
| 3. Skilled mechanical and public service workers | 17 | 14 |
| 2. and 1. Semi and unskilled workers | 1 | 1 |
| 4 to 1 combined | 42% | 37% |

[a] Information not available on all fathers.

The attempt to obtain boys in the two countries from the same backgrounds was quite successful, as Table 1 shows. As expected, the students came predominantly from the upper middle class and therefore constitute an "élite" group in each country, but the proportion of them from different backgrounds was roughly the same in both countries. The difference between Levels 6 and 5 is not significant and, in any case, is more apparent than real because there were 12 students in the German sample whose fathers were teachers in *Höhere Schule*. According to the coding system, high school teachers in the American sample were classified as minor professionals and placed in Level 5, but the fact of the matter is that such teachers in Germany have a considerably higher status than high school teachers in this country and probably belong at the very least somewhere between occupational Levels 6 and 5. If, to take account of this, half of them were classified in Level 6 or Levels 6 and 5 are combined, as in Table 1, then the distributions of backgrounds of the boys in the two samples become very nearly identical. Furthermore, the percentages of Protestants and Catholics were around 65% and 28% in both groups. In short, any nationality differences found cannot readily be attributed to differences in the socioeconomic level of the two groups of *Ss*, or to age differences (since they were matched for age), or to differences in educational plans (since they were all preparing for college), or to differences in religious background.

## RESULTS

The mean score for each of the 38 items on the questionnaire was computed for both the

German and American *Ss*.[4] About one half of the items revealed differences between the means for the two nationality groups that were significant at or beyond the .05 level. It would be easy and tempting to work out *post hoc* interpretations of these differences, but probably unwise to do so. Even though great care was taken to make the German translations as precise as possible, it is still highly likely that with emotionally toned sentiments such as those used in the questionnaire, the immediate impact of the statement would not be exactly the same on German and American audiences. So at least some of the differences might be attributed to differences in the way the sentiments were expressed. Furthermore, several of the items that on the surface looked as if they should be tapping the same sentiments yielded contradictory results. For example, the Germans agreed much more with Item 29, "There is no satisfaction in any good without a companion" ($p < .01$), but agreed less ($p < .05$) than the Americans with Item 2, "No sane, normal, decent person would ever think of hurting a close friend." Both items deal on the surface with the importance of friendship, but the Germans subscribe more to the former and the Americans to the latter, which incidentally is one of the items drawn (without the final words "or relative") from the F scale on which Germans generally score higher (4). With such variable results, how can one draw general conclusions as to which country values friendship more highly?

For these reasons it was decided to perform a factor analysis on the questionnaire to pick out those items which belonged together in each country. Such a procedure ensures that the items to be combined are chosen beforehand on the empirical basis of covariation, rather than afterwards, in order to explain the particular grouping of results on individual items. Factor analysis also tends to minimize the importance of the translation factor by yielding groups of items that must have been reacted to in somewhat similar ways in the two countries if matching factor patterns can be obtained. A correlation matrix was set up for each country[5] consisting of 36 of the 38 objective items, the two measures of motivation (n Achievement and n Affiliation), plus the number of egocentric virtues mentioned as being desirable to teach children and the total number of activities the *S* listed. Inspection of the matrices showed that practically none of the a priori clusters hung together. That is, items that had been written to tap the same sentiments were not intercorrelated consistently in both countries. So a factor analysis became all the more urgent. It was clear from inspection that certain items showed what appeared to be only random relations with other items. It was therefore decided to include in the factor analysis only those items whose average intercorrelation with the other items departed from zero at the .05 level or better, or those items whose correlations with other items showed a rate of dispersion exceeding the .05 level of probability. However, in order to make matching of factor patterns in the two countries easier, any item was included in both analyses if it met either criterion of significance in *either* country. Rotation to simple structure was performed blindly on the matrix of 19 items chosen by this method and yielded results readily interpretable in terms of two matching factors plus a set of items that shifted from one factor to the other in the two countries, as shown in Table 2. Factor A includes five items in each country that saturate $\pm .30$ or more, most of which appear to deal with striving, rationality, and planning. The two most highly saturated items are from the so called v Achievement scale (6), which reflects for the most part the extent to which a person consciously subscribes to achievement sentiments for himself. For simplicity, this factor has been named "rational striving" and perhaps even the one deviant item dealing with the respect due an older man can be thought of as belonging under this heading in the sense that the older man (father figure?) is a kind of symbol or identification figure for people with these sentiments. Factor B contains four items that saturate fairly highly in both countries.

[4] These average scores together with the German and American version of each question have been deposited with the American Documentation Institute. Order Document No. 5423 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[5] We are especially indebted to Ralph Haber who ran off the correlations on a high-speed computer.

TABLE 2

MATCHING FACTOR PATTERNS OF ITEMS FROM A QUESTIONNAIRE ADMINISTERED TO UNITED STATES AND GERMAN MALES OF HIGH SCHOOL AGE

(Saturations less than ±29 omitted)

| | Factor A | | Factor B | |
|---|---|---|---|---|
| | U.S. | Germany | U.S. | Germany |
| Factor A: Rational striving, rationality | | | | |
| 34. I work like a slave at everything I undertake until I am satisfied with the results | 43 | 47 | — | — |
| 8. I set difficult goals for myself which I attempt to reach | 52 | 35 | — | — |
| 28. Respect is due an older man no matter what kind of a person he is | 38 | 37 | — | — |
| 35. A child should never be asked to do anything unless he is told why he is asked to do it | 34 | 38 | — | — |
| 6. Nowadays with world conditions the way they are, the wise person lives for today and lets tomorrow take care of itself | −30 | −41 | — | — |
| Factor B: "That's life," resignation | | | | |
| 16. There are some people like great artists and musicians who can be forgiven for not being considerate of others, kind to the poor, etc. | — | — | 33 | 56 |
| 12. Planning only makes a person unhappy since your plans hardly ever work out anyway | — | — | 36 | 43 |
| 23. When a man is born, the success he's going to have is already in the cards, so he might as well accept it and not fight against it | — | — | 31 | 37 |
| 13. There is no such thing as a really permanent friendship. Your friends change with circumstances | — | — | 40 | (25) |
| C items: (those which shift from Factor A to Factor B) Idealistic concern for others, for "decent" social behavior | | | | |
| 27. A man with money cannot really learn how to behave in polite society if he has not had the proper upbringing | — | 48 | 47 | — |
| 2. No sane, normal, decent person would ever think of hurting a close friend | — | 38 | 39 | — |
| 7. It is better to go without something than to ask a favor | — | 30 | 31 | — |
| 18. There is hardly anything lower than a person who does not feel a great love, gratitude, and respect for his parents | — | 29 | 40 | — |
| Items with unique patterns | | | | |
| 40 Egocentric values mentioned more often than sociocentric as desirable for children | 38 | −40 | — | — |
| 22. It would irritate me very much to have a watch or clock which was off by several minutes every day or so | −32 | 46 | — | — |
| 24. My political opinion is easily swayed by editorials I read | — | −40 | — | — |
| 11. When I see a man working in an ordinary unskilled job, I often wonder why he doesn't try to do better in life | — | — | — | −52 |
| 15. If you get bad news it is better to hide what you feel and behave as if you didn't care | — | — | — | 33 |
| 38. A man who gives his life to helping others in trouble is better than a man who spends his life trying to work out schemes so that there will be less trouble in the world | — | — | — | — |

It seems to reflect a somewhat passive or resigned attitude toward life with a touch of cynicism, an attitude perhaps best summed up in a shrug of the shoulders and the French phrase "c'est la vie." The interesting thing is that four items switch from being saturated on Factor A in Germany to being saturated on Factor B in the United States. All of them seem to involve a kind of formal concern for the "proprieties," for the abstract code of politeness and obligation to others. As would be expected from this, score on the C items is correlated +.40 ($p < .05$) with Factor A score in Germany but only .06 in the United States. In contrast, the score on the C items is correlated +.28 ($p < .05$) with the score on Factor B in the United States and −.13 in Germany. In other words, explicit concern for "decency" goes with resignation and cynicism in the United States and with rational striving in Germany. The difference is nicely highlighted by the saturations for Item 40, which shows that frequent mention of egocentric virtues is negatively related to Factor A in Germany and positively related in the United States; or to put it more understandably, frequent mention of sociocentric virtues (e. g., loyalty, kindness, honesty) is positively associated with rational striving (Factor A) in Germany but negatively in the United States.

Some of the remaining items with unique factor loadings in the two countries also add to the clear-cut total picture that emerges. For example, perhaps having an inaccurate watch is irritating to Germans high in Factor A be-

cause it keeps them from being orderly in fulfilling their obligations, whereas in the United States where orderliness is not a part of the Factor A pattern, being irritable about a watch is not part of the self-picture of the "rational striver." In short, in Germany Factor A is definitely sociocentric: conscientious striving is included in a network of idealistic sentiments relating to decent social behavior. In the United States, rational striving is part of no such framework. On the contrary, sentiments of idealistic concern for "decency" are related to a kind of "un-American" passivity, resignation, and cynicism. In passing, it is perhaps worth noting that two of the C items are from the F scale (one slightly modified) and the other two have an F-scale flavor, a fact which suggests that F-scale-like sentiments may be part of a deviant, rather cynical attitude toward life in the United States, whereas in Germany they are organized into an idealistic pattern dealing with order, efficiency, and striving.

More light on the German-American contrast is shed by Table 3 which shows the mean scores for each nationality group on each of the three sets of items derived from the factor analysis, the two measures of motivation, and the two kinds of extracurricular activities in which the students said they were engaged. The Germans score significantly higher on Factor A ("rational striving") and the C items ("concern for decency") as expected, and as the first line in Table 3 shows, the difference cannot be attributed to a general tendency of the Germans to acquiesce more to items of this sort. That is, since the Germans tend to agree more with all three sets of items included in Table 3, it was necessary to check their acquiescence to all items in the questionnaire to make sure that the differences on these sets of items were not due simply to a stronger tendency for Germans to go along with or acquiesce in emotional sentiments of all sorts (2). No such tendency appeared. The German students further list themselves as being engaged in more outside activities of an individual nature—that is, in activities that do not require group organization or participation.

The Americans, on the other hand, score higher on n Achievement and much higher in the number of extracurricular *group* activities that they list. Since the difference in n Achievement score is of considerable importance in subsequent interpretation, and since it might have arisen from coding differences or the "pulling power" of the pictures used, it was carefully checked by comparing subcategory frequencies for each of the pictures used. The German students gave less achievement imagery for each of the pictures used, the smallest difference being for the picture of the student and the largest for the "father-son" picture, which may have been somewhat inappropriate for the Germans since the cattle in the background rather suggested to them an American scene. Nevertheless, even among the Americans this picture evoked the least amount of achievement imagery, and even with it omitted, the difference between the two groups is still significant. So it would be hard to attribute the results to differences in the appriateness of picture cues for the two groups.

However, it was found that the German

TABLE 3

MEAN SCORES FOR VALUES, MOTIVES, AND OUTSIDE ACTIVITIES OF U. S. AND GERMAN MALES AGED 16–19 PREPARING FOR COLLEGE

| Variable | U.S. ($N$ = 74) | German ($N$ = 87) | Diff. | $t$ | $p$ |
|---|---|---|---|---|---|
| Acquiescence score: (36 items)[a] | | | | | |
| M | 3.74 | 3.91 | −.17 | .58 | NS |
| *SD* | 1.12 | 1.33 | | | |
| Factor A: (5 items) "Rational striving" | | | | | |
| M | 15 45 | 17.30 | −1.85 | 2.40 | <.02 |
| *SD* | 4.83 | 4.81 | | | |
| Factor B: (4 items) "That's life" | | | | | |
| M | 8.36 | 9.01 | −.65 | 1.08 | NS |
| *SD* | 3.50 | 4.09 | | | |
| C items: (4 items) "Concern for decency" | | | | | |
| M | 17.92 | 20.90 | −2.98 | 4.13 | <.001 |
| *SD* | 4.88 | 4.14 | | | |
| n Achievement | | | | | |
| M | 4.69 | 2.70 | 1.99 | 2.93 | <.01 |
| *SD* | 4.45 | 4.01 | | | |
| n Affiliation | | | | | |
| M | 5.34 | 5.48 | −.14 | .24 | NS |
| *SD* | 3.70 | 3.67 | | | |
| Group activities | | | | | |
| M | 5.18 | .97 | 4.21 | 12.76 | <.001 |
| *SD* | 2.60 | 1.01 | | | |
| Individual activities | | | | | |
| M | 1.45 | 2.33 | −.88 | 3.52 | <.001 |
| *SD* | 1.78 | 1.24 | | | |

[a] Two items dropped, one because it was too complex to be understood, the other because nearly everyone in both countries agreed fully with it.

scorers had used the category, "task imagery,"[6] significantly more often than the American scorers had because, being less familiar with the achievement motive scoring system, they tended, when in any doubt at all, to use this category, whereas the American scorers had been instructed to use it only as a last resort. Consequently, the German mean score might be lower than the American because more real instances of achievement imagery had been classified as doubtful in Germany, which would also mean according to the scoring system that further subcategories could not be scored in such instances. For this reason, a random sample of 46 of the German protocols was re-scored in German in the United States by someone with considerable experience in scoring n Achievement in English. The correlation between these scores and the original scores was .70, which is lower than desirable, probably because of the difference in the use of the "task imagery" category. Still, the revised mean score for the German sample was actually *lower* (2.30) than the one given in Table 3, indicating that most of the instances of "task imagery" in the German sample should really have been classified as unrelated to achievement at all. So there seems little doubt that so far as this comparison is concerned, the German students are significantly lower in n Achievement score.[7]

The American boys, then, think more frequently in terms of doing well, a finding that is also supported by the greater frequency with which they mention egocentric rather than sociocentric virtues as desirable to teach to children. "Egocentric" is, of course, not meant in an evaluative sense, but refers to the development of individual capacities (to be intelligent, to appreciate music, to enjoy life, etc.). Unfortunately, the German coders did not force all virtues listed either into an egocentric or sociocentric category as the American coders did, but a comparison of the differences in frequency of mention of the two kinds of virtues in the two countries is still legitimate. That is, in both countries egocentric virtues were mentioned more often, but the difference in favor of them over sociocentric virtues was significantly larger ($t = 2.52$, $p < .02$) in the United States than in Germany.

If the Americans are so concerned with achievement, what keeps them in line? What prevents unbridled individualism? How are they to fulfill their obligations to society? The answer apparently lies in their much greater participation in group activities. They are not brought up on sentiments of obligations to others as the Germans are, but from kindergarten on they regularly participate in many more extracurricular functions of a group nature. In fact, by far the most impressive result in Table 3 is the low number of group activities listed by the Germans (about 1, on the average) as compared with the Americans (about 5, on the average). In these activities the American student must learn a good deal more about getting along with other people and doing things cooperatively, if these clubs are to function at all. If there is any doubt on this point, there is an interesting pair of items correlated positively both in Germany (+.20) and the United States (+.27) on which the two groups of students differ considerably. They are the following:

| Mean scores U.S. | Germany | |
|---|---|---|
| 3.30 | 2.37 | #9. The negative opinion of others often keeps me from seeing a movie or play I had planned to attend. |
| 2.84 | 2.15 | #24. My political opinion is easily swayed by editorials I read. |

Both groups tend to disagree with these items, but the Germans much more than the Americans ($t = 3.77$, $p < .001$). What is clearly suggested here is the American's greater sensitivity to public opinion, to being guided or directed by others as Riesman has so persuasively put it (15). The American's obligation to develop himself, to achieve, is checked and channelled by the opinion of others in the groups in which he participates so that individualism does not run rampant.

The German is also "individualistic" but in a somewhat different way. He subscribes more

[6] In the scoring system for n Achievement (12) "task imagery" has a value (= 0) between "unrelated" (= −1) and "achievement-related" (= +1) imagery.

[7] Ordinarily such protocols would be scored without knowledge of which group they belonged to but since they were written in different languages, this procedure was not possible in the present case without translating one set into another language. In our view, translation was more objectionable methodologically than the procedure followed here, especially since coding for n Achievement by highly trained scorers is ordinarily done with a re-score reliability of .90–.95, which does not leave much room for unconscious bias to operate.

than the American to sentiments about rational striving (Factor A) and engages in more individualistic outside activities. Furthermore he agrees more with two other items on the questionnaire which shed light on the nature of this attitude toward the self even though, as evidence, answers to the items do not rank high because they are uncorrelated in either country. The items are as follows:

| Mean scores U.S. | Germany | |
|---|---|---|
| 3.28 | 4.88 | #7. It is better to go without something than to ask a favor. |
| 3.09 | 4.92 | #15. If you get bad news, it is better to hide what you feel and behave as if you didn't care. |

In both cases the Germans agree much more with the item ($p < .001$) than the Americans do, suggesting that they have more pride in self-control and perhaps think of the self as having the will power to exercise restraint in such cases. How is this kind of self to fulfill its obligations to society? Clearly not by "learning from acting cooperatively in groups" as in America but from a strong sense of social obligation, of one's *duty* to others and of the necessity to put such an obligation above selfish interests.

Evidence was also sought for linkages between the two motives measured and the three groups of value items or the extracurricular activity patterns, but none was found. In correlating the two motive scores with the three types of value scores (Table 3) and with the total number of individual or group activities in each of the two countries (16 different correlations) only one correlation reached the .05 level of significance and it might well have been obtained by chance. The correlation was between n Affiliation and Factor A ("rational striving") in Germany ($r = -.22$). It gains added significance only because the same correlation in the United States was $+.21$ and the difference between the two correlations is highly significant. If anything, this result would appear to mean that spontaneous concern for others (n Affiliation) may make it harder to subscribe to an ideological code of individualistic achievement in Germany, whereas n Affiliation may be positively related to rational striving in the United States because achievement values there are more connected to working with others. Nevertheless, the overwhelming impression of this analysis is that motive scores (as obtained from fantasy) and value scores (as obtained from questionnaires) are not correlated and are essentially independent variables. Consequently, great care should be taken to specify the method of measurement when one speaks about such things as an "interest in achievement," which may refer either to the n Achievement score and its correlates or something like Factor A with its correlates.

Finally, an analysis was made to see if the students nominated by their peers as future outstanding citizens differed in any systematic ways from those not so nominated. For the most part they did not. As far as item or factor differences on the questionnaire are concerned, the two groups of "leaders" did not differ significantly from the "nonleader" groups within either country or from each other between countries. Only one finding of possible future interest emerged: in both countries nominated "leaders" had higher average n Affiliation than "nonleaders" ($p < .05$), a fact which suggests that among peer groups, even across nationality differences, it is the spontaneous interest of individuals in others (n Affiliation) that tends to make them "stand out" as potential leaders in the perception of the group members. Further interpretation of these results is not justified in view of the large number of insignificant findings and the possibility that such a "nomination" procedure may not actually select "leaders" in a very stable or meaningful way.

## DISCUSSION

The findings of the present study may be most simply discussed in terms of the following table which has been organized in terms of the double obligation of "the individual in society" to fulfill some of his own needs and also to satisfy some of society's needs:

| | Obligation to self | Obligation to society |
|---|---|---|
| United States | High n Achievement, concern for egocentric, "self-development" virtues | High participation in group activities, greater sensitivity to the opinions of others |
| Germany | High v Achievement (concern for rational striving), greater participation in individualistic activities, more emphasis on power of self to exercise restraint | Greater concern over obligations to an idealistic code of decency (governing interpersonal behavior) |

Presumably every culture develops more or less patterned solutions to these two problems if it exists for long, and the individual member learns these norms and reflects them in his behavior and attitudes. The point to notice is that according to this analytic scheme, *both* obligations must be at least in part fulfilled by a majority of its members if the society is to function at all. If society should demand too much of its members, it should eventually collapse through *apathy* or *rebellion* because its members can no longer satisfy their needs. On the other hand, if society is neglected through exclusive concern with individual satisfaction, it should eventually collapse through *anarchy*. No one is interested in the welfare of others and society degenerates into the "war of all against all." Between these two theoretical extremes, there should be all sorts of ways of structuring the two sets of obligations which produce working social systems of greater or lesser efficiency. The data of the present study suggest that the scheme is of some value in presenting the somewhat different ways in which morality is structured in the United States and Germany. In the United States a high spontaneous interest in achievement is counterbalanced by much experience in group activities in which the individual learns to channel his achievement needs according to the opinions of others. In Germany a strong sense of self as a striving, controlling entity is offset by an equally strong sense of obligation to a *code* of decency. Interestingly enough, the American "value formula" appears to be largely unconscious or informally understood, as compared to the German one, at any rate. That is, the American boys have higher n Achievement, which is *uncorrelated* with a conscious value placed on striving, and they participate in many group activities without, it is fairly safe to say, having a very explicitly formulated *moral* justification for such behavior. Yet we are arguing that functionally group activities have just such a moral basis in checking excessive individualism. For the Germans, on the other hand, the whole matter appears to be much more consciously worked out. The moral code is more formal and ideological, as other students of German character have observed (14). Individuals are taught and "know" they have a great obligation to the group and strong egos, which are, however, capable of making sacrifices of personal feelings to meet this obligation. In the present study, the German value pattern appears most clearly in the sentiments expressed on the questionnaire, while the United States value pattern appears most clearly in the projective material and the kind of activity the students engage in.

The present study is obviously of very limited scope. It is based on small samples of German and American boys of predominantly upper middle-class status. But the generalizations drawn from it suggest a much wider application to the cultures involved. What evidence is there from previous studies that our conclusions do or do not apply more generally? Has anything been added or subtracted to what is already known about German and American "national character"? In the main, our results are supported by what others have found previously, with one possible change in emphasis to be noted below. Probably the most sensitive clinical or anthropological comparison of the two countries was made by the social psychologist, Kurt Lewin (11), shortly after he immigrated to the United States from Germany. His observations agree at almost every point with our conclusions. He found "probably no other people as interested in *individual accomplishment* . . . as the Americans" (11, p. 277). On the other hand, he calls attention to the flexibility of Americans, to the ease with which they meet strangers, or appear to change their "personality" (opinions and attitudes) to adjust to new situations. In short, he recognized what Riesman has called the *other-directedness* (15) of many Americans which clearly appears in the data of this study.

So far as the Germans are concerned, he described them as carrying more of their "specific individual characteristics to every situation," as having a greater sense of their own identity as separate individuals than Americans do. This kind of "individualism" we have also met in the reports of our German students with the additional note that the self-concept carries with it connotations of personal striving and self-control. Lewin sees the German's sense of obligation to society largely in terms of his tendency to obey orders

from higher authority without question. He says, for instance, "In Germany it seems to be the natural right of the adult to rule and the duty of the child to obey" (11, p. 270). In this respect his opinion is supported by a large number of research workers (1, 3, 4, 8, 13, 16). Their point of view is perhaps best summarized by McGranahan as follows: "The German is noted for his unquestioning obedience to authority, his failure to exercise individual responsibility and act on the basis of independent moral judgment" (13, p. 248).

It is at this point that our study differs most from previous ones, and even here the difference is not so much one of fact, but of interpretation. There seems to be no reason to doubt the evidence that Germans are more obedient. The problem arises when attempts are made to explain why they are so obedient, even when such obedience causes great personal suffering or pain to others. Explanations have frequently been given in terms of such motivating factors as "identification with the aggressor," "authoritarian-submission," "sado-masochistic trends," and the like. All such explanations suffer from doing little more than naming the behavior (i.e., Germans obey because they *like to submit*) and from interpreting it in ways which, however acceptable they may be to outsiders (particularly if they dislike Germans), do not make sense to Germans as the rationalizations that at least consciously lead them to obey. The present study does not pretend to get at unconscious factors behind German behavior, but it does shed some light on why they behave as they do *from their point of view*.

To point up the issue as sharply as possible, compare the average agreement with the following item by our United States and German boys:

| Mean scores U.S. | Germany | |
|---|---|---|
| 3.50 | 5.12 | #35. A child should never be asked to do anything unless he is told why he is asked to do it. |

German boys want explanations significantly more ($p < .01$) than American boys do. On the surface at least, this finding would appear to contradict the notion that Germans are noted for "unquestioning obedience," but one must distinguish between behavior and the attitudes behind it. Suppose we assume that other investigators are correct, that German boys do obey more promptly than American boys, then our problem is to understand why they demand explanations for orders. Here lies the key to the difference between our interpretation and some others. We believe they obey not because they in some sense like obedience and therefore "obey blindly" but because they have some reason, preferably some idealistic reason, for obeying. Without the reason, without the ideal, obedience is meaningless. Hence, we believe, the German emphasis in our data on obligation to an idealistic code of decency.

To make the contrasting interpretations even clearer, consider the following item from McGranahan's study (13) of German and American high school students:

> Do you think a boy is justified in running away from home if his father is cruel and brutal?

Three out of four of the American boys said they felt the boy was justified in running away, whereas the German boys split about 50-50. McGranahan is inclined to interpret this finding as supporting the hypothesis that Germans are more submissive and more used to obeying family authority without question no matter how cruel it may be. But our previous analysis suggests that the item does not properly distinguish between what Germans do and why they do it. They may in fact be more submissive to cruel fathers, but not, if our interpretation is correct, because it is good per se to submit but because it is a way of discharging one's duty to an ideology of proper behavior to others (including parents as a very special case). In other words, the German boy may understand the question about running away from a cruel father in terms of whether he is willing to endure personal discomfort to follow the code governing children's obligations to their parents, which is seen by him to be clearly correct in the abstract so that not living up to it for personal reasons seems like a form of selfishness. In fact, the overriding importance of this code of social obligation is frequently stressed in German literature by pointing out that it not only may cause suffering to one's self (thus overriding selfishness) but to others as well. McGranahan and Wayne in an interpretation very similar to ours note that in German plays

"we find . . . various murders of beloved persons consciously carried out by sympathetically portrayed characters" (14, p. 454). The characters are sympathetic because they *must* sacrifice even their "nearest and dearest" in the interests of a "higher," more idealistic goal. The same theme appears in a popular German fairly tale, *Der treue Johannes*, in which a father must cut off the heads of his twin sons to fulfill an obligation to a servant whom he has wronged. Apparently, Germans learn in this way that the ultimate test of the sincerity of one's commitment to a code of idealistic obligation is the extreme to which one will go in the sacrifice of personal interests or even the interests of *particular* others. It is in this latter sense that the German concern with decency to others is "ideological" rather than "personal," as McGranahan and Wayne (14) point out.

Outsiders may well wonder how a people so concerned ideologically with what we have called "decency" were also capable of so much cruelty to others under the Nazis. Many have struggled with this problem at deep psychodynamic levels, but a clue to its solution may lie in the two characteristics of the code of obligation just mentioned. First, it seems to be ideological rather than personal. It applies to "others" in the abstract, not to "particular others," so that it doesn't seem so inconsistent to sacrifice particular others in the interest of an abstract code of right conduct designed for the welfare of the whole group. It is perhaps worth noting in passing that the negative correlation in Germany between n Affiliation and Factor A (sentiments of rational striving and decency) may reflect the conflict between the "ideological" and the personal. It may be harder to subscribe to an *ideology* of personal striving and decency if one has much spontaneous interest in *particular* others (n Affiliation) in a country where particularistic interests are often thought of as conflicting with one's larger obligations.

Second, cruelty may actually appear as the validating sign of the strength of one's commitment to the code, as in *Der treue Johannes*. One must show that he can "take" even doing horrible things to *prove*[8] his commitment to a higher end. Since on this point McGranahan and Wayne's study came to the same general conclusion based on pre-Nazi data as ours does on post-Nazi data, it appears that what the Nazis did was to play up a tendency to idealistic sacrifice of personal interests, even of affiliative ones, always present in German character.

In the main then, our conclusions are supported by other students of German and American character. We differ chiefly in seeing German obedience as a response that at least from the internal point of view is dictated not by a liking for obedience but by a strong belief in the importance of sticking to an abstract, ideological code of right conduct in society. As confirmation of this alternative interpretation, McGranahan and Wayne (14) found that obedience is *not* always good in German plays: but the one time rebellion is legitimate is when it is in the interest of some higher good of society as a whole. In short, it is not obedience per se which is good. It is seen as instrumental to an ideology of right social behavior that may on occasion demand rebellion.

## Summary

A questionnaire covering various value sentiments and a picture-story test for measuring n Achievement and n Affiliation were administered to 87 German and 74 United States male students preparing for college who had been carefully matched for age (around 18 years) and socioeconomic background (largely upper middle class). A factor analysis of the questionnaire results yielded two factors that could be matched in the two countries and a third set of items that shifted from one factor to the other across countries. The "switch-over" items which dealt largely with decency and obligation to others were associated with Factor A ("rational striving") in Germany and were more strongly accepted by the Germans than by the Americans, as were also the Factor A items. In the United States, the switch-over items were associated with Factor B ("resigned cynicism") suggesting that acceptance of statements about strong obli-

[8] One may even wonder, somewhat paradoxically, if the Germans are forced, functionally speaking, to emphasize self-sacrifice so much just because they have such a strong sense of self. Perhaps the problem is less acute for the American whose sense of personal identity is less well-developed in the sense that he talks more freely to others (Items #7, #15) and is more easily guided by others (Items #9, #24).

gations to others tend to be associated in the United States with a kind of deviant "un-American" passive resignation. Further analysis showed that the American students were significantly higher than the Germans on n Achievement, on participation in group activities, and in the extent to which they said they were guided by the opinions of others. The Germans for their part reported a greater number of individualistic outside activities and subscribed more to sentiments picturing the ego as independent of others. These results were interpreted to mean that in the United States obligation to self appears as a strong unconscious need for Achievement, and obligation to society as participation in many group activities, and that in Germany obligation to the self appears as a stress on the ego as a separate, unique, willing entity and obligation to society as a moral imperative to place obligation to the "impersonal other" as an abstract code above particularistic, "selfish" considerations. An attempt is made to interpret German "authoritarianism" in terms of duty to such an abstract code rather than in terms of the more usual, somewhat unanalytical and ethnocentric assumption that Germans "like" to dominate and submit more than other people.

## REFERENCES


1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritian personality*. New York: Harper, 1949.
2. BASS, B. M. Authoritarianism of acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
3. BRICKNER, R. *Is Germany incurable?* Philadelphia: J. B. Lippincott, 1943.
4. COHN, T. S., & KARSCH, H. Administration of the F-scale to a sample of Germans. *J. abnorm. soc. Psychol.*, 1954, **49**, 471.
5. COMMAGER, H. S. *America in perspective*. New York: Random House, 1947.
6. DECHARMS, R., MORRISON, H. W., REITMAN, W. R., & MCCLELLAND, D. C. Behavioral correlates of directly and indirectly measured achievement motivation. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955.
7. DE TOCQUEVILLE, A. *Democracy in America*. (1833). New York: Knopf, 1945.
8. FROMM, E. Hitler and the Nazi authoritarian character structure. In T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947.
9. GORER, G. *The American people*. New York: Norton, 1948.
10. KLUCKHOHN, F. R. Dominant and substitute profiles of cultural orientations: Their significance for the analysis of social stratification. *Social Forces*, 1950, **28**, 376–393.
11. LEWIN, K. Some socio-psychological differences between the United States and Germany. *Charact. and Pers.*, 1936, **4**, 265–293.
12. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
13. MCGRANAHAN, D. V. A comparison of social attitudes among American and German youth. *J. abnorm. soc. Psychol.*, 1946, **41**, 245–257.
14. MCGRANAHAN, D. V., & WAYNE, I. German and American traits reflected in popular drama. *Hum. Relat.*, 1948, **1**, 429–455.
15. RIESMAN, D., WITH GLAZER, N., & DENNEY, R. *The lonely crowd*. New York: Doubleday, 1953.
16. SCHAFFNER, B. *Fatherland: A study of authoritarianism in the German family*. New York: Columbia Univer. Press, 1948.
17. SHIPLEY T. E., JR., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.


Received April 29, 1957.

# SEX DIFFERENCES IN PROBLEM-SOLVING PERFORMANCE AS A FUNCTION OF ATTITUDE DIFFERENCES[1]

GLORIA L. CAREY[2]
*University of Nebraska*

PROBLEM solving is one of several areas of intellectual functioning in which sex differences in performance have been noted (1, 3, 5, 6). Usually these differences have been reported as incidental to the major findings, and very few studies have been designed specifically to test sex differences in reasoning or problem-solving ability. Sweeney (7), however, did perform a series of experiments in which he controlled variables he thought might be responsible for the previously noted masculine superiority in problem solving. Equating men and women on aptitude (mathematical and verbal), visualization, mechanical, comprehension or mathematical background did eliminate sex differences in performance on certain types of problems, but a striking number of highly significant sex differences remained. Sweeney concluded that these remaining differences occurred on problems that required restructuring or involved a general reasoning factor.

The present investigation began with the idea that some of the sex differences in problem solving could be accounted for by differences in nonintellectual factors. The question that needed to be asked is: How well do women like to solve problems? If problem solving is an activity preferred by men and disliked by women, there exists a good possibility that sex differences in performance are actually a reflection of sex differences in attitude.[3]

Such observations as the relative lack of women in the scientific professions, feminine expressions of inadequacy and dislike in problem-solving experiments, and the cultural stereotype of *man* as the "reasoning animal" point toward a negative attitude toward problem solving and the reasoning process, and make the following hypothesis plausible: sex differences in problem-solving performance which are not the result of differences in general intelligence, special aptitudes, or information, are attributable to differences in attitude toward problem solving.

In order to test this hypothesis, studies were designed to answer the following specific questions:

1. Can a scale be constructed that measures attitude toward problem solving?
2. Is there a sex difference on such a scale?
3. Is problem-solving attitude related to problem-solving performance?
4. Is an attempt to change problem-solving attitude experimentally followed by a change in performance?
5. Do women respond more favorably than men to an attempt to improve their attitudes? If women have an initially more unfavorable attitude than do men, there should be more opportunity for improving the attitude of women than of men, and hence one might expect a greater relative improvement in their performance.

Affirmative answers to the first three questions would give support to the major hypothesis, but it was felt that such answers to the fourth and fifth questions would provide even more convincing evidence. In addition, a scale throws little light on the question of whether performance depends on attitude, or whether the reverse is also true—that attitude depends on performance. Therefore, the main experiment was designed to answer Questions 4 and 5, which concern a deliberate manipulation of the independent variable (attitude) and measurement of its effect on the dependent variable (performance).

[1] Condensed from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Stanford University, 1955. The writer is indebted to Drs. D. W. Taylor, P. R. Farnsworth, R. R. Sears, and Olga McNemar for their help and advice. She also wishes to thank Drs. Yvonne W. Brackbill, D. L. LaBerge, and R. J. Rhine for their assistance in conducting experimental sessions. The investigation was carried out under Project NR 150-149 supported by Contract N6onr 25125 between Stanford University and the Office of Naval Research. Permission is granted for reproduction, translation, publication, use, and disposal of this article in whole or in part by the United States Government.

[2] Now at Swarthmore College.

[3] By "attitude" the writer is referring simply to liking or not liking to solve problems, or to the valence which problem solving has for the individual. This does not mean that it is possible to distinguish clearly between attitude and other nonintellectual factors in problem-solving success, such as motivation and interest.

## METHOD

*Attitude scale.* The Likert scaling method was applied to 63 statements which, in the opinion of several investigators working on a problem-solving research project, might be related to attitude toward problem solving. (Example: "I would rather have someone tell me the solution to a difficult problem than to have to work it out for myself.") There was no attempt deliberately to construct items which would show a sex difference in response. On the basis of an item analysis of the responses of 32 men and 100 women in an elementary psychology course at Stanford University, 36 items were selected that differentiated between high and low scorers at the .05 level or better. The 36 items were then split on the basis of similarity of content and form into two groups of 18 items each, which were referred to as Forms A and B of the attitude scale. For this first sample, the intercorrelation between the two forms administered at the same time was .94. The intercorrelation for a second sample of 59 women and 50 men was .83.

*Attitude modification.* In order to study the hypothesized effect of attitude on performance, the following experiment was designed so as to measure attitude and performance both before and after an attempt was made to modify attitude. Sixteen groups of six subjects (*S*s) each met together for a total of three hours, one hour on each of three successive days. Each group was composed of three men and three women—making a total of 48 men and 48 women. All *S*s were members of an elementary psychology class at Stanford University, and the majority of them were sophomores. During the first hour *S*s were given one form of the attitude scale and also a series of ten problems to solve, with 40 minutes working time allowed for completion of the problems. The problems were ones for which a significant sex difference had appeared in an earlier experiment (7) even though the sexes had been matched on variables thought to be related to performance differences. The second hour consisted of an informal group discussion of some of the factors involved in problem-solving success, with the implicit aim of (*a*) conveying an appreciation of the relationship between attitude toward problem solving and problem-solving success and (*b*) encouraging each *S* to explore and recognize his or her own attitude toward problem solving. During the third hour, the alternate form of the attitude scale and a second, comparable set of ten problems were administered. Since the two sets of problems were not exactly equal in difficulty, the order of presentation was reversed for half the *S*s. At no time was a reference made to the real purpose of the experiment, nor was the subject of sex differences ever introduced. As far as the *S*s were concerned, the first and third hours of the experiment were used to standardize the problems and scales, and the second (discussion) hour was used to get their suggestions for future problem-solving research.

There were four experimenters (*E*s), two male and two female, each of whom worked with four groups of six *S*s each. They administered the attitude scales and problems, and led the group discussions. They also observed two demonstration discussions by the writer before leading discussions of their own. The use of four *E*s allowed more generality to the conclusions than would be possible with one *E*. Two *E*s of each sex also provided an opportunity to test whether a significant interaction occurs between sex of *E* and sex of *S*.

If a relatively greater improvement in performance were obtained for women than for men, the possibility would exist that women benefited more from the practice involved in working the first set of problems than did men. As a group, women may have had less experience with the types of problems used and hence may show a greater practice effect. For this reason, control groups of 28 women and 20 men were used, the procedure being the same as that for the experimental *S*s, with the exception that the discussion period was eliminated for the control groups. In each control group, the order of administration of the problems and attitude scales was reversed for half the sample, as it was for the experimental *S*s.

## RESULTS

The results obtained from the attitude scale and from the attitude modification experiment are summarized with respect to each of the five questions raised earlier.

1. *Can a scale be constructed that measures attitude toward problem solving?* A reliable measuring instrument was constructed. Some internal evidence of validity was provided by the fact that the items that referred directly to interest in or frequency of problem-solving activity were among the most highly differentiating items in the item analysis. Additional evidence of validity was provided by the fact that attitude scores and performance scores were positively related (see Question 3), which indicates that the people who say they like to solve problems or are interested in related activities also solve the most problems. It would be desirable to know the frequency with which *S*s with high attitude scores actually seek or indulge in problem-solving behavior, but even without such a behavioral criterion of validity the contents of the scale and measures of validity cited above provide some evidence that the scale is actually measuring attitude toward problem solving.

2. *Is there a sex difference on such a scale?* Men received significantly higher scores on the attitude scale than did women. In Table 1 is a summary of the data for the two samples used.

3. *Is problem-solving attitude related to problem-solving performance?* Correlations were computed between the attitude scale scores and problem-solving scores for men and women separately in both the prediscussion and the postdiscussion testing sessions. The eight correlations, seven of which were in the predicted direction, are given in Table 2. The fact that

TABLE 1
SEX DIFFERENCES IN ATTITUDE

| Sample | N | Mean | SD | t | p |
|---|---|---|---|---|---|
| First Sample | | | | | |
| Form A | | | | | |
| Men | 32 | 63.12 | 8.83 | 3.08 | .01* |
| Women | 100 | 57.61 | 8.71 | | |
| Form B | | | | | |
| Men | 32 | 58.19 | 9.68 | 2.74 | .01* |
| Women | 100 | 53.17 | 8.70 | | |
| Second Sample | | | | | |
| Form A | | | | | |
| Men | 50 | 55.88 | 7.56 | 1.86 | .04* |
| Women | 59 | 52.68 | 9.86 | | |
| Form B | | | | | |
| Men | 50 | 60.08 | 7.10 | 2.39 | .01* |
| Women | 59 | 56.42 | 8.48 | | |

* An asterisk in this and the following tables indicates that a one-tailed test of significance was made, since the direction of the difference was predicted.

TABLE 2
CORRELATIONS BETWEEN ATTITUDE AND PERFORMANCE SCORES: EXPERIMENTAL GROUPS

| Session and Sex | N | Attitude A and Problems I | | Attitude B and Problems II | |
|---|---|---|---|---|---|
| | | r | p | r | p |
| Prediscussion | | | | | |
| Men | 24 | .57 | .001* | .24 | .13* |
| Women | 24 | .29 | .08* | −.08 | — |
| Postdiscussion | | | | | |
| Men | 24 | .60 | .001* | .32 | .07* |
| Women | 24 | .20 | .18* | .29 | .08* |

only two were highly significant may have been due in part to the small $N$ for each subgroup.

4. *Is an attempt to change problem-solving attitude experimentally followed by a change in performance?* Performance data for the pre- and postdiscussion testing sessions are given in Table 3. A significant improvement in the performance of women indicates that the question may be answered in the affirmative for the female *S*s.

5. *Do women respond more favorably than men to an attempt to improve their attitudes?* In order to answer this question, a two-way analysis of variance design was used to analyze the difference scores, with classification according to *E* and sex of *S*.[4] Table 4 is a summary of the

[4] As a check on the consistency of the improvement or lack of improvement for the women in the groups led by the four *E*s, Bartlett's test for homogeneity of variance was applied to the variance estimates of the four *E*s' groups. Since chi square = 4.49 ($p$ = .19), the assumption of homogeneity of variance was not rejected. That is to say, the groups did not differ in variability.

TABLE 3
DIFFERENCES IN PROBLEM-SOLVING SCORES BETWEEN FIRST AND SECOND SESSIONS: EXPERIMENTAL GROUPS

| Sex | N | First Session | | Second Session | | Diff. | t | p |
|---|---|---|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD | | | |
| Men | 48 | 5.13 | 2.04 | 5.33 | 1.99 | .20 | .48 | — |
| Women | 48 | 3.78 | 1.81 | 4.63 | 1.73 | .85 | 2.30 | .02* |

TABLE 4
ANALYSIS OF VARIANCE OF PROBLEM-SOLVING DIFFERENCE SCORES: EXPERIMENTAL GROUPS

| Source | Sum of Squares | df | Var. Est. | F | p |
|---|---|---|---|---|---|
| Sex difference | 10.24 | 1 | 10.14 | 2.86 | .05* |
| Exper. difference | 40.45 | 3 | 13.48 | 3.80 | .03 |
| Interaction | 8.44 | 3 | 2.81 | .79 | — |
| Individual differences | 312.51 | 88 | 3.55 | | |

analysis of variance results. Difference scores were calculated by subtracting each *S*'s problem-solving score for the prediscussion session from that for the postdiscussion session. Since the relative improvement of female *S*s (mean difference score of women minus mean difference score of men) was 1.25 and 1.22 for one female *E* and one male *E*, respectively; and −.08 and .22 for the other female *E* and male *E*, respectively, there was obviously no significant interaction between sex of *E* and sex of *S*. Hence, the classification according to sex of *E* was not used.

The sex difference in amount of improvement in performance was significant, women responding more favorably than men to the attempt to improve problem-solving attitude. An additional finding was that the sex difference in amount of improvement varied according to *E*, a fact that indicates the need for using several *E*s to yield results of some generality.

The change in performance of the control groups was not significant for either sex ($t$ = .77 and .28 for men and women, respectively). This finding indicates that the amount of practice represented by the first session did not result in a significant improvement in performance.

As shown in Table 5, the improvement of the

TABLE 5
DIFFERENCES BETWEEN MEAN DIFFERENCE SCORES FOR EXPERIMENTAL AND CONTROL GROUPS

| Sex | Mean Difference | | Diff. | $t$ | $p$ |
|---|---|---|---|---|---|
| | Experimental | Control | | | |
| Men | .20 | .55 | −.35 | .78 | — |
| Women | .85 | −.14 | .99 | 2.11 | .02* |
| Difference between men and women | | | 1.34 | 2.06 | .02* |

experimental as compared with the control groups was significantly larger for women than for men. This finding further confirms the conclusion reached earlier on the basis of the analysis in Table 4: women showed significantly more improvement as a result of the discussion than did men.

A comparison was also made of the pre- and postdiscussion attitude scores. Contrary to expectation, women did not show a relatively greater change in attitude scores than did men, nor did the scores improve significantly for either sex following the discussion. Male and female *S*s in the control groups also showed no significant improvement in attitude scores and a comparison of the change for experimental versus control *S*s revealed no significant differences.

## DISCUSSION

The answers to the five questions support the hypothesis that previously reported sex differences in problem-solving preformance are attributable in part to sex differences in attitude toward problem solving. An explanation of this difference in attitude probably lies in our culture, which still adheres to some extent to the view that "Woman is the unreasoning animal and pokes the fire from on top."[5] It is hypothesized that men tend to be more favorably disposed toward problem solving because they have been encouraged in this attitude and because acknowledgment of an unfavorable attitude is a denial of one aspect of the masculine role. Some support for this view is shown in the relationships found between attitude toward problem solving and mathematical and verbal aptitude (2). The correlations for women's aptitude and attitude scores were highly significant for both aptitude measures, but there was no relationship between men's aptitude and attitude scores—a finding which indicates that men tend not to admit to a dislike for problem solving even when they are lacking in intellectual aptitude.

Although a sex difference in improvement of problem-solving performance was the major prediction, it was also predicted that there would be a sex difference in improvement of attitude scores. The lack of any improvement in attitude scores warrants further consideration.

One explanation for this finding is that the scale was not sensitive enough to reflect the changes in attitude that occurred, since a number of the items referred to interests and activities with respect to which *S*s would have had no opportunity to alter their behavior before the second testing session (e.g., "I enjoy being the score keeper when playing bridge or canasta"). Another explanation for the lack of improvement in attitude scores is based on the occurrence of catharsis during the discussion. In this permissive situation *S*s were encouraged to explore their attitudes toward problem solving—whether positive or negative—with the result that they felt quite free to express negative attitudes. The cathartic effect of discussion may have been such as to result in a freer, more realistic expression of attitude.

A final possibility, of course, is that the discussion did not modify attitude and that other factors accounted for the improvement in women's performance. Practice in solving the first set of problems might result in such an improvement if women were more affected by the practice than were men, but this factor was controlled by the use of female *S*s who had a similar exposure to the problems without the intervening discussion. The fact that the control *S*s did not improve in performance suggests that the effect of solving the first set of problems was to counterbalance the positive effects of practice by reinforcement of the existing negative attitudes. An examination of the discussion procedure indicates that motivations such as sex rivalry, increased level of aspiration, or a desire to please the discussion leader do not appear to account for the improvement in women's performance. At no time did *S*s in the discussion groups comment on sex differences in problem-solving performance, and the *E*s were under instructions not to refer to such differences or to the real purpose of the discus-

[5] Quoted by William James, in a discussion of why women are relatively poor at reasoning (4, p. 368).

sion. In fact, *S*s did not know that they would be given more problems to solve after the discussion session, although the remark was occasionally made at the end of the discussion that if *S*s had a chance to work on more problems, they thought they would do a better job of solving them. No scores or answers were given out during the experiment so that change in level of aspiration is also not a very likely possibility. By the use of both male and female *E*s, interaction between sex of *E* and sex of *S* could be determined, but the results indicated that such interaction did not occur. Since *E*s made no attempt to exhort *S*s to improve their attitudes or performance, it is unlikely that female *S*s improved in performance because they were motivated to please *E*, although this possibility cannot be ruled out entirely.

Consideration of possible alternative explanations for the significantly greater improvement in women's performance leaves change in attitude as a tenable one even though the attitude scale itself was probably not sensitive enough to reflect a change.

A final comment may be made on the fact that there were marked differences among *E*s in the relative improvement of female *S*s. The *E* difference was not due to an interaction between sex of *E* and sex of *S*, and the discussion leaders were of approximately the same age and level of academic attainment. Since the discussion procedure was also fairly well standardized from one group leader to another, it would appear that the differences were related to personality characteristics of the *E*s.

## SUMMARY

The study was designed to test the hypothesis that sex differences in problem-solving performance which are not the result of differences in general intelligence, special aptitudes, or information, are attributable to differences in attitude toward problem solving. First, a scale designed to measure attitude toward problem solving was constructed. Then the attitude scale and a series of ten problems were administered to experimental groups consisting of three men and three women each. The group members then participated in a discussion, with the disguised intent of promoting a more favorable attitude toward problem solving. In a postdiscussion session, *S*s were administered an alternate form of the attitude scale and a second set of problems, comparable to those used in the prediscussion session. A control group that did not participate in the discussion was also used.

The data were analyzed according to their bearing on five questions related to the major hypothesis. 1. Can a scale be constructed that measures attitude toward problem solving? 2. Is there a sex difference on such a scale? 3. Is problem-solving attitude related to problem-solving performance? 4. Is an attempt to change problem-solving attitude experimentally followed by a change in performance? 5. Do women respond more favorably than men to an attempt to improve their attitudes?

The attitude scale gave some internal evidence of validity, and men were found to have significantly higher scores on the scale than women. Also, attitude scores were found to have some positive relationship to performance scores. After the group discussion, designed to induce a more favorable attitude toward problem solving, women showed a significant improvement in performance, while men showed no change. An examination of some possible ways of accounting for the relatively greater improvement of women's postdiscussion performance indicated that improvement of attitude is at least a tenable one. In general, the findings support the hypothesis of sex differences in problem solving as a function of sex differences in attitude. An explanation of such attitudinal differences in terms of cultural differences for men and women is offered.

## REFERENCES

1. BILLINGS, M. L. Problem solving in different fields of endeavor. *Amer. J. Psychol.*, 1934, **46**, 259–272.
2. CAREY, GLORIA L. *Reduction of sex differences in problem solving by improvement of attitude through group discussion.* Stanford: Dept. Psychol., Stanford Univer., Tech. Rep. No. 9, 1955 (Contract N6onr 25125).
3. GUETZKOW, H. An analysis of the operation of set in problem-solving behavior. *J. gen. Psychol.*, 1951, **45**, 219–244.
4. JAMES, W. *Principles of psychology.* Vol. 2. New York: Henry Holt, 1890.
5. MAIER, N. R. F. Reasoning in humans. III. The mechanisms of equivalent stimuli and of reasoning. *J. exp. Psychol.*, 1945, **35**, 349–360.
6. SAUGSTAD, P. Incidental memory and problem solving. *Univer. Chicago Psychometr. Lab. Rep.*, 1951, No. 66.
7. SWEENEY, E. J. *Sex differences in problem solving.* Stanford: Dept. Psychol., Stanford Univer., Tech. Rep. No. 1, 1953 (Contract N6onr 25125).

Received April 29, 1957.

# PERSONALITY AND IQ CHANGE[1]

JEROME KAGAN, LESTER W. SONTAG, CHARLES T. BAKER, AND VIRGINIA L. NELSON

*The Fels Research Institute, Antioch College*

RESEARCH on mental development during the last twenty years has indicated that a child's IQ score does not necessarily remain constant with age (2, 3, 4, 10). Several reports (9, 10, 12) suggest that changes in environmental conditions can depress or raise IQ level and it is sometimes implied that these changes may be explained by recourse to personality variables. The purpose of this paper is to demonstrate that changes in IQ during childhood are correlated with certain personality predispositions as inferred from projective test data. The personality variables under study include (*a*) need for achievement, (*b*) competitive strivings, (*c*) curiosity about nature, and (*d*) passivity.

Performance on an IQ test is assumed to be a function of at least two major variables; the variety of skills and abilities the person brings to the test situation and his motivation to perform well on the test (2, 6). Since the IQ scores of some children change markedly during the school years, it seems plausible to assume that those children who show marked increases in IQ have a very strong motivation to acquire or develop the various intellectual skills tapped by an IQ test and to perform well in a testing situation. It is suggested that need for achievement, competitive strivings, and curiosity about nature motivate the acquisition and improvement of cognitive abilities and by so doing facilitate increases in tested IQ.

The social environment often awards praise and recognition for intellectual accomplishment, and school age children with a high need for achievement might seek to gratify this need through intellectual activity. Thus it was predicted that children showing marked increases in IQ would produce more achievement imagery on the TAT than those with minimal gains in IQ.

Secondly, the school environment emphasizes competitive intellectual activity, and children with strong competitive needs would be highly motivated to acquire the intellectual skills which result in successful competition with one's classmates. Thus it was predicted that children showing IQ gains would show more competitive strivings than children displaying minimal gains in IQ. In choosing an index of competitive strivings, besides the related measure of TAT achievement fantasy, it was decided to use aggressive content on the Rorschach. The bases for this choice rested on the assumptions that (*a*) incidence of aggressive imagery reflected degree of aggressive motivation and (*b*) competition was a socially accepted form of aggressive behavior. For in competition, as in aggression, the child desires to defeat another individual and assert his superiority over him. The population of children in this study is predominantly middle class and apt to place strong inhibitions on direct, overt expression of aggression. Therefore, there would be a tendency for the individual with high aggressive motivation to seek socially accepted channels for aggressive expression such as competitive activity with peers. Thus it was predicted that children showing IQ gain would report more Rorschach aggressive content than those with minimal gain because of their greater competitive predisposition.

A third motive that might facilitate a child's acquisition of knowledge and skills in dealing with the environment could be curiosity about nature. Interest in birth, death, sexual anatomy, and other processes of nature is a frequent phenomenon in young children. It is suggested that the more intense this curiosity the greater the motivation to acquire the habits which would gratify this motive. Since reading, questioning, and manipulating the environment are effective behavioral methods of gratifying one's curiosity, it might be expected that the highly curious child would be more likely to develop these skills and therefore apt to gain in IQ score. The TAT measure used to evaluate curiosity was presence of themes of

[1] This investigation was supported in part by a research grant (PHS M 1260) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The writers wish to thank Dr. Seymour B. Sarason for his critical reading of the manuscript.

interest in nature and its phenomena. For the Rorschach, it was hypothesized that concern with the body might reflect, in part, heightened interest in natural processes, and it was suggested that anatomy content might be more frequent for children who showed marked IQ gains than for those with minimal increases in IQ. It is recognized that many clinical psychologists regard anatomy content in adults as indicative of psychopathology. This study is concerned with the correlates of IQ gain rather than psychopathology, and it is not implied that children who show increases in IQ are completely free of conflict. Secondly, it was felt that the determinants of anatomy content for children might be different from those which produce this content in adults.

A final prediction dealt with the predisposition to behavioral passivity. The children who show IQ gains have been characterized as having high need achievement, competitive strivings, and curiosity about the environment. This constellation of motives implies that when these children are confronted with a problem, they would have a tendency to attack and attempt to solve the problem rather than withdraw from the situation or seek help. On this basis, it was predicted that children who showed IQ gains would be less likely than those with minimal IQ increases to characterize their TAT heroes as passive in attitude or behavior.

The Fels Research Institute is uniquely equipped to test these ideas about IQ change since it has continuous longitudinal information on the development of a sample of normal children. These data include intelligence and projective tests, observations of the children, and reports on the parent-child interaction. In a recent study, Sontag, Baker, and Nelson (11) related personality information on a sample of children with changes in IQ and found that those children who showed marked increases in IQ were rated as more competitive, more likely to display self-initiated behavior and less passive than those who showed decreases in IQ. The TAT and Rorschach protocols were not utilized in making these personality ratings, and the results from this study served as a major stimulus for the present investigation.

## METHOD

A sample of 140 Fels subjects (*Ss*), 70 of each sex, were chosen for study because a fairly complete record of test information was available on them. From ages 2½ to 6, the Stanford-Binet intelligence test (1916 or 1937 revision) was administered to most *Ss* twice yearly, on their birthdays and six months after their birthdays. From ages 6 to 12, most *Ss* received alternately Form L or Form M of the 1937 revision annually on or near each *S*'s birthday. All of the tests were administered by one of the authors (VLN). The mean IQ of the Fels population is near 120, with standard deviation varying from 14 to 20 IQ points.

In order to obtain groups of *Ss* who showed the most change in IQ score from ages 6 to 10, a smoothed longitudinal plot of each *S*'s IQ was prepared by averaging the mean of three consecutive test scores around each age. This procedure is explained in detail in other reports (1, 10, 11). This technique tends to eliminate erratic variations in IQ and hopefully furnishes a more valid measure of IQ changes. Then each *S*'s smoothed IQ at age 6 was subtracted from his smoothed IQ at age 10, and this distribution of differences, positive if *S* gained in IQ and negative if *S* lost in IQ, was divided into quartiles. This report deals with the projective test information on those *Ss* in the two extreme groups; those who increased and those who decreased the most in IQ score. These will be called Group A, the IQ ascenders, and Group B, the IQ descenders, respectively. There was no significant difference between the mean IQ of the two extreme quartiles at age six, the means being 119 and 116 for Groups A and D respectively. The average amount of increase in IQ for Group A was larger (plus 17 points) than the corresponding decrease for the members of Group D (minus 5 points) and while 46 per cent of Group D lost five or more points, every child in Group A gained 10 or more points during the years 6 through 10. The mean IQ of the entire sample of 140 tends to increase slightly from ages 6 to 10, probably as a result of practice effects with the same test. Since every *S* in Group D showed a decrease in IQ, it might be inferred that the members of Group D did not benefit from practice and familiarity with the test, and it is probably more accurate to view Group D *Ss* in this light rather than as *Ss* who showed marked decreases in IQ score.

The projective tests used in the analysis were the Rorschach and selected TAT pictures. Two factors governed the choice of the TAT cards which were analyzed. Because the protocols were gathered over a period of years, there was not complete comparability for all *Ss* for the number of cards administered. Secondly, the specific hypotheses of the study dictated the cards chosen for analysis and Cards 1, 3 BM, 3 GF, 5, 6 BM, 12 F, 14, and 17 BM were selected for analysis. The age at which the TAT protocols were administered ranged from 8–9 to 14–6 with median at 11–6 and 80 per cent of the protocols obtained between the ages of 11 and 12. The age at which the Rorschachs were administered ranged from 6–5 to 13–6 with median at 10–5 and 63 per cent of the sample having had the test between ages 10 and 11. Since the Rorschach and TAT were administered by different examiners there was no comparability with respect to inquiry or probing. Thus, the analysis of both the Rorschach and TAT was restricted to the *S*'s spontaneous verbalization to the stimulus before any questions or inquiry were conducted by the examiner. The protocols were scored for the following fantasy categories.

1. *Need achievement on the TAT.* Achievement imagery on the TAT was scored according to the definition of McClelland et al. (8); and themes involving a reference to competition with a standard of excellence were scored achievement imagery.

2. *Rorschach aggression.* The definition of aggressive content on the Rorschach included (*a*) people, animals, or creatures engaged in physical or verbal aggression, e.g., fighting or quarreling, (*b*) explosive objects or explosions, e.g., volcanoes, bombs exploding, fireworks, and (*c*) objects or animal parts normally regarded as instruments of aggression, e.g., spears, rifles, clubs, guns, knives, horns, and claws.

3. *Intellectual curiosity about nature.* For the TAT, curiosity was defined in terms of themes in which someone is interested in the processes or phenomena of nature. Curiosity on the Rorschach was restricted to anatomy or X-ray responses of internal organs or boney parts, e.g., stomach, backbone, ribs.

4. *Passivity.* Because of the limited amount of thematic material in the spontaneous performance, themes of passivity were limited to stories in which the central figure was described as sleepy, tired, or resting.

The fantasy categories were independently scored by the senior author and an assistant without knowledge of the *S*'s IQ scores.[2] Reliability was very high because of the limited amount of content scored for each response and the objectivity of the definitions. Percentage of agreement for the three TAT categories was 95 per cent and for the two Rorschach categories 99 per cent.

## RESULTS

Although there was a total of 70 *S*s in the two extreme quartiles, not all of the *S*s had Rorschach or TAT data for the age range under study. Table 1 shows the distribution of *S*s, by sex and direction of IQ change, for the TAT and Rorschach analyses. Because there are approximately twice as many boys as there are girls in Group A, all comparisons were first made separately by sex and results were only combined if the direction of the result for both boys and girls in the same IQ group was in the predicted direction.

1. *Need achievement.* All achievement themes, save one, occurred to Cards 1 and 17 BM. The typical achievement story to Card 1 concerned a boy who wanted to master the violin and/or become a famous violinist, while the typical achievement theme to 17 BM involved competitive activity with regard to rope climbing. Table 2 shows the percentage of *S*s in each group reporting achievement imagery plots to Cards 1, 17 BM, and to both pictures.

For both Cards 1 and 17 BM, more male and female *S*s in Group A report achievement imagery than the boys or girls of Group D. For Card 1, the difference between Group A and Group D girls is reliable at the .03 level; the difference for boys is in the predicted direction but not significant. For Card 17 BM, the difference between Group A and Group D boys is significant ($P = .03$) and in the predicted direction for girls. All $P$ values are for one tail and were evaluated using the exact method suggested by Fisher (5). When the sexes were pooled, comparisons between Groups A and D were significant not only for Cards 1 and 17 BM separately but also for the number of *S*s telling achievement imagery to both Cards 1 and 17 BM ($P < .10$, .03, and .01 respectively). Thus, the *S*s who showed increases in IQ were more prone to structure Cards 1 and 17 BM in terms of achievement oriented behavior than the *S*s in Group D.

2. *Aggressive content on Rorschach.* There was no significant difference between Groups A and D or between boys and girls with respect to the mean number of responses per protocol, and the mean for the entire sample was 27 responses. There was no difference between Group A and Group D girls with respect to percentage of each group reporting one or more aggressive responses per protocol (30.0 per cent for Group A versus 33.0 per cent for Group D). However, the difference between Group A and D boys approached significance with 59.1 per

TABLE 1

DISTRIBUTION OF *S*s BY SEX AND DIRECTION OF IQ CHANGE USED IN THE ANALYSIS OF THE TAT AND RORSCHACH

| Group | TAT | | Rorschach | |
|---|---|---|---|---|
| | Boys | Girls | Boys | Girls |
| Group A | 22 | 11 | 22 | 10 |
| Group D | 10 | 20 | 9 | 18 |
| Both groups | 32 | 31 | 31 | 28 |

TABLE 2

PERCENTAGE OF *S*s REPORTING ACHIEVEMENT IMAGERY TO CARDS 1 AND 17 BM

| TAT Card | Group A | | | Group D | | |
|---|---|---|---|---|---|---|
| | Boys | Girls | Boys and Girls | Boys | Girls | Boys and Girls |
| Card 1 | 36.4 | 50.0 | 40.6 | 27.3 | 15.0 | 19.4 |
| Card 17 BM | 36.4 | 30.0 | 34.4 | 0.0 | 15.0 | 9.7 |
| Cards 1 and 17 BM | 22.7 | 10.0 | 18.8 | 0.0 | 0.0 | 0.0 |

[2] The writers wish to thank Mary Schnurer for her assistance in assessing the reliability of the scoring.

cent of the former and 22.2 per cent of the latter reporting one or more aggressive images ($P = .07$). Thus, the prediction of a correlation between IQ increase and aggressive imagery held only for the boys. Because of the tentativeness of this result and the more speculative nature of the hypothesis relating competitive striving and aggressive content, an attempt was made to validate this finding by analyzing a later Rorschach protocol for the boys in Groups A and D. Not all of the boys had Rorschachs administered to them at a later age, and only 15 *S*s in Group A and five in Group D were available for analysis. The median ages at the time of administration were 13-8 and 15-0 for Groups A and D respectively, and there was no significant difference in the lengths of the protocols of the two groups. The results were in the same direction for 86.7 per cent of Group A, and 20.0 per cent of Group D reported one or more aggressive images, and this difference is highly significant ($P = .01$).

3. *Intellectual curiosity.* The only TAT card eliciting curiosity plots was Card 14, and the typical theme described a person gazing at or interested in the stars or the heavens. Table 3 shows the percentage of each group telling such themes to Card 14.

Both the boys and girls in Group A told more themes of interest in the stars or heavens than the males and females in Group D ($P = .14$, $P = .10$, respectively) and combining of the sexes yielded a highly significant difference between Groups A and D ($P < .01$).

4. *Anatomy and X-ray responses on the Rorschach.* There was no difference between Group A and Group D girls reporting one or more anatomy responses (30.0 per cent versus 38.9 per cent for Groups A and D respectively). For the boys, 31.8 per cent of Group A and 0.0 per cent of Group D reported anatomy or X-ray imagery, a difference that approached significance ($P = .06$). This finding was also validated on the same sample of 20 boys that was used to check the differences in aggressive content. The results were in the same direction with 60.0 per cent of Group A and 20.0 per cent of Group D reporting anatomy content ($P = .15$).

5. *Passivity.* Card 3 BM accounted for most of the passivity themes and the groups were compared with respect to the incidence of stories to Card 3 BM in which the central figure was sleepy, tired, or resting. Table 4 shows the percentage of each group telling such themes. Both the boys and girls in Group D showed more passivity themes than the boys and girls in Group A. Although only the difference for the girls was significant ($P = .06$), when the sexes were pooled the difference was highly reliable ($P < .03$).

Cards 3 GF, 5, 6 BM, and 12 F did not furnish data relevant to the hypotheses under test and these results are not summarized.

TABLE 3

PERCENTAGE OF *S*s REPORTING THEMES OF CURIOSITY TO CARD 14

| Sex | Group A | Group D |
|---|---|---|
| Boys | 40.9 | 18.2 |
| Girls | 30.0 | 5.0 |
| Boys and girls | 37.5 | 9.7 |

TABLE 4

PERCENTAGE OF *S*s REPORTING THEMES OF PASSIVITY TO CARD 3 BM

| Sex | Group A | Group D |
|---|---|---|
| Boys | 9.1 | 27.3 |
| Girls | 10.0 | 45.0 |
| Boys and girls | 9.4 | 38.7 |

## DISCUSSION

In the main, the hypotheses about the differences between Groups A and D have been verified. Boy and girl ascenders produced more TAT achievement imagery and curiosity about nature than Group D children and male ascenders displayed more aggressive content on the Rorschach than the boys in Group D. The higher incidence of aggressive imagery for the boys who gained in IQ was interpreted as reflecting stronger competitive motivation. Finally, the *S*s in Group D were presumed to have a more passive orientation since they were more likely to perceive the ambiguous figure on Card 3 BM as sleeping or tired. The relation between Rorschach anatomy content and IQ gain was the most tentative finding.

The results are interpreted as indicating that high motivation to achieve, competitive strivings, and curiosity about nature may motivate the acquisition of intellectual skills and knowledge which, in turn, facilitates increases in tested IQ. If one accepts the generally assumed notion that boys are more competitive

and achievement oriented than girls, the fact that there were twice as many boys in Group A as there were girls supports the present interpretation. A recent study using the Edwards Personal Preference Schedule found that high school boys obtained higher need achievement scores than high school girls (7).

These results are not interpreted as indicating that strong achievement, competitive, and curiosity motives are the only variables involved in producing gains in IQ. The *Ss* in this study are all average or above in IQ and there is not adequate sampling of children with lower IQ levels. One would not expect *Ss* with low IQs or language handicaps to suddenly show an interest in reading despite achievement needs or intellectual curiosity. The child who spends increased time reading because of a heightened interest in natural processes must have already learned the basic reading skills so that this behavior is not a difficult or unlikely choice for him.

Similarly, needs for achievement and successful competition should only motivate attempts at improvement of intellectual abilities in a social milieu where praise, recognition, and superior status are awarded for such accomplishment. That is, achievement-oriented children from homes in which intellectual activity was praised would probably be more likely to master intellectual skills than achievement-oriented children from homes in which such accomplishment was not rewarded. In a cultural environment where athletic ability, fighting prowess, or success with the opposite sex was highly valued, one might expect the child to choose these behavioral channels to gratify his achievement and competitive needs. The parents in the Fels population are predominantly middle class and tend to place importance on intellectual accomplishment. A large majority of the parents have attended college, and since enrollment in the Fels program is voluntary it might be inferred that only parents who valued knowledge and scientific pursuits would be predisposed to become part of the research population. Thus, the children under study tend to come from homes which value intellectual ability.

Study of the educational attainment of the parents of the *Ss* in Groups A and D revealed no significant difference between the groups with respect to the percentage of families in which both parents attended college (57.1 per cent for Group A versus 42.9 per cent for Group D; $P > .30$). Although there is a slight difference favoring the educational level of Group A families, the difference was not dramatic. There may be important differences between Groups A and D with respect to the differential encouragement of intellectual achievement, but measurement of these differences would probably require variables more refined than educational level of the parents. However, even though parental emphasis on intellectual activity may increase the child's desire to improve his cognitive skills, the child's predisposition to adopt or rebel against parental values should selectively influence his motivation to strive for intellectual accomplishment. Thus, the type of relation between parent and child may be an important factor in this process.

Finally, there is the possibility that genetic and/or constitutional variables may play a role in facilitating marked IQ changes. There is considerable data indicating that genetic factors influence general IQ level but less evidence relevant to the role of these variables in producing childhood increases in IQ score. For most of the children in our population, IQs tend to level off during the ages 6–10 and most of the marked changes in level occur during the preschool years. However, the exact relationship between genetic variables and IQ change has yet to be determined. The phenomenon of IQ increase during the school years is admittedly complex and it is not implied that the child's motives are the major factor. However, it is suggested that personality needs may influence this process. Perhaps the most accurate generalization is that for middle-class children with average or above IQ levels, strong achievement, competitive, and curiosity needs may facilitate IQ gains by motivating the child to master intellectual skills.

A final implication of these findings is that they add indirect evidence for the usefulness of the Rorschach and TAT as research instruments. Validation of a predicted relationship between TAT achievement imagery and IQ gain increases one's confidence in the hypothesis that TAT plots can serve as an index of achievement-oriented tendencies. The results of the Rorschach analysis suggest that aggressive content may be an index of an individual's aggressive predispositions but not necessarily a measure of his tendency to express direct,

physical aggression. Although Sontag, Baker, and Nelson (11), using behavioral observations, rated the boys in Group A as more competitive than those in Group D, there was no difference between these groups with respect to intensity or incidence of direct verbal or physical aggression or destruction of property. We have assumed that competition is a socially approved form of aggressive behavior and the higher incidence of aggressive content for Group A boys was presumed to be a result of their more intense competitive strivings. Some clinicians who use projective tests are too prone to focus on predictive statements about direct, physical aggression when confronted with a protocol containing aggressive content. One is apt to overlook the fact that the individual may have alternative behavioral channels for expression of aggressive motives.

## SUMMARY

For a group of 140 boys and girls in the Fels Research population on whom continuous Binet IQ data were available, a distribution of IQ change was obtained by subtracting each *S*'s smoothed IQ at age 6 from his smoothed IQ at age 10. This distribution of differences was divided into quartiles, and the Rorschach and TAT protocols of the upper, (maximum increase in IQ) and lower (maximum decrease in IQ) quartiles were analyzed and compared. The results showed that in comparing the *S*s who showed IQ increases with those showing IQ decreases, the former had, on the TAT, significantly more (*a*) achievement imagery on Cards 1 and 17 BM and (*b*) themes of curiosity about nature on Card 14, and significantly fewer themes of passivity on Card 3 BM. For the boys only, more of the *S*s who increased in IQ had anatomy responses and aggressive imagery on the Rorschach. The results were interpreted as indicating that high need achievement, competitive striving, and curiosity about nature are correlated with gains in IQ score because they may facilitate the acquisition of skills that are measured by the intelligence test.

## REFERENCES

1. BAKER, C. T., SONTAG, L. W., & NELSON, VIRGINIA L. Specific ability in IQ change. *J. consult. Psychol.*, 1955, **19**, 307–310.
2. BAYLEY, NANCY. Mental growth in young children. *Yearb. Nat. Soc. Stud. Educ.*, 1940, **39**, (II), 11–47.
3. BAYLEY, NANCY. Consistency and variability in the growth in IQ from birth to eighteen years. *J. genet. Psychol.*, 1949, **75**, 165–196.
4. BRADWAY, KATHERINE. IQ constancy on the Revised Stanford-Binet from the preschool to the junior high school level. *J. genet. Psychol.*, 1944, **65**, 197–217.
5. FISHER, R. A. *Statistical methods for research workers*. (5th ed.) Edinburgh: Oliver & Boyd, 1934.
6. HAGGARD, E. A., DAVIS, A., & HAVIGHURST, R. J. Some factors which influence performance of children on intelligence tests. *Amer. Psychol.*, 1948, **3**, 265–266.
7. KLETT, C. J. Performance of high school students on the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1957, **21**, 68–72.
8. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
9. RICHARDS, T. W. Mental test performance as a reflection of the child's current life situation: A methodological study. *Child Develpm.*, 1951, **22**, 221–233.
10. SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Personality as a determinant of performance. *Amer. J. Orthopsychiat.*, 1955, **25**, 555–562.
11. SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Mental growth and personality development. *Monogr. Soc. Res. Child Develpm.*, in press.
12. WELLMANN, BETH L., & MCCANDLESS, B. R. Factors associated with Binet IQ changes of preschool children. *Psychol. Monogr.*, 1946, **60**, No. 2 (Whole No. 278).

Received April 29, 1957.

# THE PSYCHOLOGY OF SMOKING[1]

CHARLES McARTHUR, ELLEN WALDRON, AND JOHN DICKINSON
*Harvard University*

WHO smokes? Who doesn't smoke? Who smokes too much? Who can stop smoking? Reviewing the sparse literature on smoking and personality, we could find neither descriptive data answering these questions nor good theory from which the answers might be deduced. Therefore, we have made a first attempt at delineating a psychology of smoking.

Our basic data come from The Study of Adult Development, once known as the Grant Study. That study has been described elsewhere (6). Our subjects were a panel of 252 Harvard alumni who were selected during their sophomore years for lack of visible abnormality. During their sophomore years, which fell between 1938 and 1942, they were studied by a great range of medical, physiological, psychological, anthropological, and sociological techniques. We have been following these men since by annual questionnaire, retesting, and visits. Questions about smoking habits were routinely asked in the questionnaires during the fifteen or so years the men have been followed. This very full material permitted the study staff to make comparisons between smoking habits and hundreds of variables. Only the findings relevant to psychosocial theory are reported here.[2]

## WHO DOESN'T SMOKE?

A striking number of our subjects did not smoke. When the men were first studied, as sophomores, 108 of them (some 45 per cent) had not begun to smoke at all. Sixty-one never have. That is about 24 per cent of the group. In any one year, more than 40 per cent of the men will be "off" smoking, though for some this is very temporary.

Our theory begins with Bales' very general propositions about the origins of any compulsive habit. Bales (2) began by studying compulsive drinkers. He found that "analysis shows the craving to be a result of at least two types of underlying elements: (1) some need or complex of needs. . . and (2) an orientating structure of habitual thought patterns and associated emotional justifications." Orientation has to come first; if the experience is found gratifying, it may be put into the service of needs. As Bales (3) says, "it is during the formation of the compulsive habit that the influence of the society and its culture come to a critical focus in the individual person and work their effect." In particular, it is society that determines whether the new habit seems to be "a means of relieving. . . inner tensions, or whether such a thought arouses a strong counteranxiety."

Convergent theories of anthropologists, sociologists, and psychologists, reviewed and empirically confirmed elsewhere (10, 13, 14,15), all suggest that the use of tobacco to resolve major personal needs might "arouse a strong counteranxiety" in upwardly mobile members of America's lower middle class. It seems likely that this group differs in their attitude toward smoking from people who stand both above and below them in the social hierarchy. Their attitude to smoking would, presumably, be corollary of a basic value system that these people hold in contrast to most of contemporary society.

The source of this value system is described by Weber (23) as the Protestant Ethic. It is a work morality, emphasizing the Devil's stake in idleness and self-indulgence. It is an ethic that abhors as Sin the wasting of one's substance in unprofitable trivia, since such waste, in the original protestant dogmas, was evidence that a man had not been vouchsafed Grace. This ethic was, as Mills (18) remarks, a morality of producers, who abhorred consumers. It is typical that, after one especially long

[1] This study was supported by the Tobacco Industry Research Committee. It was conducted as part of the research program of Harvard's University Health Service.

[2] Some of the other findings of the study staff, especially those in relation to medical and physical variables, are reported in Heath, C. Differences between Smokers and Non-Smokers, in press. A few items appear in both papers, as they seem relevant.

A table, giving data for each of the findings reported, has been deposited with the American Documentation Institute. Order Document No. 5464, remitting $1.25 for 35 mm. microfilm or $1.25 for 6 by 8 in. photocopies.

Calvinistic debate, tobacco was settled upon as a good thing because the growing of it produced profit.

This theology of "provident purposes" was soon secularized in the form of "worldly asceticism," as Weber illustrates from the diaries of Benjamin Franklin. Indeed, the process of lay morality descending from religious dogma is still going on, even among the Harvard group we are studying, as Allport (1) has shown in his study of religion in the post-war college student. It is thus that the traditional middle-class values were generated.

These values are now being replaced in our society, according to sociologists. They have become the property, though perhaps not exclusively, of the "Old Middle Class." They belong to Riesman's (20) "Inner-Directed" group, so soon to be replaced by the Outer-Directed; to Mills' (18) "cheerful robots," so soon to become equally cheerful mass consumers. Yet it is these values that have generated what Mead (16) calls "the American core culture" and what Kluckhohn (10) describes as the dominant value profile in American culture and expects to find focussed in the middle class, and which McArthur (13, 14) has empirically shown to be focussed in the lower middle. It still gives rise in the American middle class to a new generation instilled with the "achievement mores" whose operation in a contemporary urban setting is described by McClelland (15). Children reared in this tradition are mobile, following what Miller and Form (17) describe as Ambitious Careers. As one study (14) shows, "The Future, Doing-oriented family must produce sons reared in the 'achievement mores,' taught to look forward to by-passing or surpassing their fathers' occupational roles.... It is the hopes of the mother that these sons must realize in order to feel successful. They... will have introjected her precepts. It is these boys who will, after college, be expected to leave the family and 'make their own way.'" Children reared in this tradition are still (15) weaned and trained early, encouraged to be ambitious, hard-working, and clean-minded. "I know that Andy has a clean mind," wrote one father of a nonsmoker, "he does not smoke, he does not loiter on the street with gangs, he has been brought up to look for the fine things in life and only by hard work and perseverance can they be achieved and enjoyed."

In the study, we have found (13) that a simple empirical tool for bringing out middle vs. upper class contrasts is to compare Harvard men who prepared for college in public high schools with Harvard men who prepared for college in private preparatory schools. The high-school graduate who goes to Harvard has usually received a heavy dose of "work morality" during his rearing. As an earlier study (14) suggested, "His role in the family often is that of standard-bearer, the son who is to win status or ethnic mobility." The prep-school boy who goes to Harvard is more likely to have been bred on a code of acceptable behavior. Tradition sets bounds on him in terms of "what is done," not in terms of "worldly asceticism" as a means of "getting ahead." If a concentration of nonsmokers exists in our data, then we may well predict that it will focus among our public-school graduates.

This is the case. In a steady progression with status, 20 per cent of the graduates of very exclusive private schools do not smoke, 30 per cent of the graduates of less exclusive private schools, and 40 per cent of the graduates from public schools. This pattern is significant at the .01 level. Part of the explanation of this fact may be that the prep-school boys learned to smoke in their dormitories (rules notwithstanding) while the high-school graduate did not experience dormitory life with its peer-orientation until coming to college. Some facts fit this notion. In retrospect, about three-quarters of the private-school graduates who smoke say that they learned to smoke in prep school or at the time of prep school; about three-quarters of the public-school graduate smokers say they learned to smoke in college. As one progresses from private boarding to private day to public day schools, there is a significant (.01) decline in the proportion of boys who already smoked at the time they were taken into our study. However, there is a status difference operating together with this dormitory factor. If we hold boarding constant, there remains a significant (.02) difference between the more exclusive and the less exclusive schools. Such a pattern may well have been generated in the families who elect to and are able to send their sons to these schools, rather than in the schools themselves. Besides, the tacit other half of the dormitory argument is that high schools

do not orient their pupils toward smoking. This cannot be said. Hollingshead (8) tells us that in "Elmtown" high school, "practically all boys (91 per cent) smoke, irrespective of age or (social) class." Our public-school nonsmokers are therefore somehow special. It may be supposed that their rearing for mobility, in terms of the old ethic described earlier, may have created a considerable "strong counteranxiety" to undermine the orientation from their peer culture toward learning to smoke.

> In Elmtown,
> Law and the mores deny high-school students the right to enjoy the pleasures derived from tobacco, gambling, and alcohol. However, the mystery with which adults surround these areas of behavior lends them a special value which seems to act as a stimulus to many young people who desire to experience the supposed thrill of pleasures their elders deny them. The conspiracy of silence which is an essential part of the clandestine violation of the mores has already taught them how easy it is to avoid restrictions imposed by law and taboo if they are discreet about how, where, and under what circumstances it is done. Acquisition of knowledge of the means of transgressing against alcohol, tobacco, and gambling taboos without being caught and the thrill of violating these taboos take place for the most part in the clique.

Presumably any peer-oriented youngster is likely to smoke, whatever his social status, as part of a kind of defiant claim on adult status. It may be supposed that the more earnest students in high school take adult prohibitions more seriously and, perhaps, do not always fit well with "the clique." Theirs may be a more compliant claim on adulthood. Perhaps among the earnest nonsmoking nine per cent of Elmtown seniors, there was one who applied to Harvard. More went to other colleges and were mobile into white collar technical and professional occupations. It is not to be supposed that attendance at any one university is a potent cause. Census data (5) show a nationwide increase of nonsmokers when professional and technical people are compared with other occupational groups. Nonsmoking in America is a social class phenomenon.

It is in England, too. A Hulton survey (7) suggested that nonsmoking was slightly commoner among men of the English middle class while heavy smoking was commoner among English working-class men. An excellent market survey[3] in England not only confirms this difference in rate of smoking but also documents in many ways the fact that nonsmoking is widely perceived as a mark of "middle-class respectability."

Our data certainly suggest that smoking is popularly defined as faintly disreputable or, at least, one of the small vices. Nonsmokers are significantly often (.01) nondrinkers; indeed, it seems that, as Straus and Bacon (22) report from Yale, drinking precedes smoking as the student takes up adult foibles. Our nonsmokers are also significantly often (.01) people who do not drink coffee.

We may expect some orientation toward nonsmoking to have been taught these men by their religions. The kind of piety that Weber describes should spill over into self-denial, even in the absence of a specific theological proscription. Proscription of smoking sometimes does occur. Among Protestant denominations, it is usually those of less social standing (19)—e.g., Fundamentalists—who condemn smoking, while those of more fashionable status —e.g., Episcopalians—do not. The lower-middle-class focus of nonsmoking may be thus overdetermined. At any rate, we may find overlap among smoking habits, social status, and piety.

Earlier in the course of the study, Heath had set up ratings of the devoutness of the participants and also of their families. These ratings were based on considerable information about theological beliefs, church attendance and activities, personal use of prayer, etc. (No attention was paid to vices, great or small, and hence smoking in itself did not enter this rating.) The boys who did not smoke when they entered the study had more devout parents (.05). They were also more likely (.01) to attend church while at Harvard, despite the absence of any chapel requirement. Lifelong nonsmoking is associated even more strongly (.02) with devoutness of parents. It would seem that the individual lastingly introjects his smoking morality! Indeed, the nonsmokers are often rated higher on devoutness than their parents.

This introjected piety seems to be more an individual matter than we expected. In the

[3] Davis, F. Cigarette Smoking Motivation Study. This is an unpublished monograph prepared by Research Services Limited, to whom we are grateful for access to it.

study data, nonsmoking is the modal pattern for the members of no one religious denomination. Nor did the interaction between piety and status work quite as expected. It is true that, as expected, most Protestant families with low incomes were rated high on devoutness and produced sons who were nonsmokers. Presumably, we see here the introjection of pietistic standards, at least in the form of "worldly asceticism," in the absence of theological proscription.

The nonsmokers may represent survivals not only of the protestant ethic and middle-class morality but also of the related phenomenon Riesman (20) calls "Inner Direction." For example, in answering the Strong Vocational Interest Blank, more sophomore nonsmokers (.03) said they preferred "nights at home" to "nights away from home," and smokers said the opposite. Also, nonsmokers preferred (.02) "belonging to few societies," smokers "belonging to many societies." These trends were especially marked among nonsmokers from public schools, whose earnest individualism is thus reaffirmed, but private-school nonsmokers showed some of the same trend.

A curious sidelight may be interpreted variously: more sophomore smokers (.01) say they would prefer "preparing the advertising for the machine," in the appropriate Strong item, while nonsmokers prefer not to. It is also true (.05) that more smokers responded "Like" to the occupation of Advertiser, nonsmokers responding "Dislike." Perhaps the nonsmoker is so Inner Directed that he does not have his "radar" tuned to the latest fad in mass consumption. He is, perhaps, not yet part of the consumer ethic that is said (18) to be appearing now among the middle class.

A related finding, based on Strong-like items in a recent questionnaire, is that smokers respond "Like" to the occupation Sales Manager (.01) but "Dislike" to the occupation Scientific Research Worker (.01). This attitude is reflected in career choices: the nonsmokers contribute more physical scientists (.01). These facts overlap the status differences already observed; few private-school boys go into science but many go into business. Whether the ethos of the physical scientist overdetermines his nonsmoking over and above the effect of his middle-class origins and upward mobility is not clear from our data. Certainly the two—mobility and science—when combined pull powerfully: of the twelve public-school boys from low-income families who became scientists, ten never smoked and one tried it but stopped.

In summary, the nonsmoker seems to have been oriented by the mores of a particular American subculture. He is often of lower-middle-class origin and himself upwardly mobile. He shows the "worldly asceticism" that has stemmed from the old Protestant ethic. Often he is pious, perhaps more so than his parents. It seems likely that he has reacted to smoking as being one of the "small vices" to which the flesh is heir. He is, at any rate, an Inner-Directed person, introjecting the morals of his youth, perhaps a serious sort, and maybe an introvert. He does not go along with the suggestions for a consumer morality offered by the mass media. He approves scientific rather than business values and may often himself be a scientist or engineer. Just what causes underlie this web of correlations is not clear, but existing theory about the ethos of the Old Middle Class or about the American core culture seems relevant. The standards this man has introjected furnish sufficiently "strong counteranxiety" to prevent his sharing the orientation toward smoking that seems to be common to all the rest of American society both below and above him.

## WHO SMOKES HEAVILY?

Whether a man smokes or no seems best explained by his social orientation. Whether he becomes a heavy smoker seems best explained by his personal needs. Bales (2) speaks of the effectiveness, once a man has been oriented toward adopting a habit, of "the degree to which the culture operates to bring about acute needs for adjustment, or inner tensions, in its members. There are many of these: culturally-produced anxiety, guilt, conflict, suppressed aggression, and sexual tensions of various sorts may be taken as examples."

There is no doubt that our very few really heavy smokers, who meet the criterion for Heavy Smoking now customary in medical research—"two packs a day for several years"—have more than their share of "acute needs for adjustment." Most have had marital problems, some quite dramatic. All are given

to impulsive acts, some to physical violence, if only in the form of volunteering for dangerous missions. Several are hard-driving, tough competitors. None are usual for our group. As one observer phrased it, "they are men who live in overdrive!" Their stories are told by Heath.[4]

That the need to smoke heavily might be generated by anxiety is, of course, a commonsense hypothesis. Our data warns us to view it with caution. We have not been able to correlate changes in smoking with changes in the tenseness of a man's life situation, in spite of the "clinical hunch" of staff members that this relationship held. Perhaps anxious smoking is episodic, so that our yearly questionnaire is too coarse a measure of it. Perhaps something is to be learned from the answers to a question about the symptoms of stress that our participants felt when under pressure. Half of the smokers said they "smoked more" in these circumstances. The interesting fact, though, is that 70 per cent of the heavy smokers but only 30 per cent of the light smokers said this. Is smoking, then, a suitable tension reducer only after it has become a firmly ingrained habit? One thinks of learning theory: a response must be "high in the habit hierarchy" before it has the "availability" to reduce tension. Or one may think that this little item is a rather pretty confirmation of Bales' contention that only after thorough orientation can a compulsive habit be used to satisfy anxious needs.

The most striking correlation between heavy smoking and personality is found in our inkblot test material. In the early days of the study, Dr. Wells (24, 25) used an abbreviated Rorschach procedure, allowing one minute for response to each of the standard blots, which he called the Timed Rorschach. The resulting scores are not identical with Rorschach scores but seem to be interpretable by similar theory. The Rorschach variable we concentrated on is the Experience Balance, defined as the ratio of the number of human movement responses to the weighted sum of the various kinds of color responses. If both members of this ratio are equal to or less than 2:2, one speaks of a "coartated" Rorschach, by which one means an emotionally narrow, drily factual, excessively controlled performance. If either member of the ratio exceeds 2, one sort or another of emotional expression has taken place, and the label "coartated" no longer applies.

Heavy smokers produce more (.05) coartated records. (Our definition of Heavy Smoking is now loosened to include men with lifetime averages of a pack a day, so as to include a workable number of cases.) What is more, coartated people tend (.01) to increase the amount of their smoking strikingly as the years go by. This is not a universal trend; many of our men reach a smoking pleateau early or vacillate between nonsmoking and moderate smoking. The coartated person is common among that plurality of our people whose smoking curve "snowballs," showing a positive acceleration. The appearance of these curves suggests to the eye that they be interpreted as showing accelerating habituation. Interestingly, such curves often appear among people who started smoking late. (They are not just "catching up," however; they soon surpass the smoking rates of men who started earlier.)

This association between accelerating smoking rates and coartation is strong enough to override the other trends so far reported. A coartated man tends to become a heavy smoker no matter what his social background or how pious his family—or himself. His orientation may show in his being a nonsmoker in college if he came from a nonsmoking type of background. Soon after college, he typically begins to smoke, and the habit accelerates. One is reminded of Bales' (2) comment that, "there is reason to believe that if the inner tensions are sufficiently acute, certain individuals will become compulsively habituated in spite of opposed social attitudes."

Are our coartated people that badly off? There is no very detailed Rorschach theory (9, 21) from which one might interpret their tests. They do seem to indicate a certain lack of emotional resources—or unwillingness to use them. Understandably, these men were rated by the psychiatrists (26) as inarticulate, pragmatic, and bland. They were not necessarily thought to have inner tensions. They were, at any rate, quite hard to get to know. Many were rated "Just-So" by the psychiatrist; that is, they were compulsive fiddlers, desk-

[4] Heath, C. Differences between Smokers and Non-Smokers, in press.

arrangers, people who allayed their tensions through fussy activities, of which smoking could understandably become one. This very striking finding about coartation remains hard to explain. Nor have we found the means to cross-validate it. Since the college Rorschach pattern predicts lifetime smoking, another long-term study would be needed for ideal cross-validation.

Not all heavy smokers are coartated, of course. Among the noncoartated subjects, there is a tendency for the associations between amount of smoking and some of the psychosocial variables previously discussed to be particularly "clean." It is as though coartation washes out the array of emotional variations between persons as well as, perhaps, the variability of the smoking rate. Among these noncoartated subjects, it does appear that the psychosocial variables have more to do with the difference between nonsmoking and moderate smoking (as discussed above) than with heavy smoking.

In the noncoartated group, moderate smoking and heavy smoking are respectively related (.001) to the psychiatrist's rating of Strong Basic Personality and Weak Basic Personality. One presumes that the latter rating is appropriate to the rather uncontrolled men who were described as being among our very heaviest smokers. It is not true that everyone who smoked more than a pack a day over his adult life was poorly integrated. It is true that poorly integrated people were much more commonly found among these heavier smokers. Again, one sees a pattern such as Bales would predict: orientation leading to moderate smoking but "some need or complex of needs for adjustment" leading to excessive habituation.

In summary, then, we may hypothesize that starting to smoke is largely brought about by one's social environment but that reactions to smoking, once it has started, seem to depend in good part on the personal needs that the newly-established habit is able to gratify. Some people seize on the habit compulsively. These people may often be emotionally constricted types for whom there is great gain in a simple "flight into behavior" or they may be restless, active men, for whom smoking is just one more impulsive activity. It would also seem that anxious people can seize on smoking as a tension reducer if they have already, for other reasons, been oriented toward it. In short, the habit, once well available, increases in strength if it serves well the person's emotional economy.

### WHO CAN STOP SMOKING?

As every habitual smoker knows to his sorrow, ability to quit or cut down decreases as smoking increases. If we divide our lighter from our heavier smokers at an adult lifetime average of half a pack a day, the contingency relation between "lighter" and "heavier" vs. "can stop" and "can't stop," as shown in recent responses to questionnaires, gives a $p$ less than .001. Even more striking is the march (.01) of the mean numbers of cigarettes smoked per day during adult life. For men who can stop smoking this figure (computed only during their smoking years) is 9, for men who don't try to stop it is 18, and for those who cannot stop it is 20. The sheer amount of tobacco so far consumed is by far the largest difference between the group of men who can stop and the group who, at least so far, cannot.

The variables that were related to becoming a heavy smoker seem also to bear some relation to ability to stop. Thus, coartated Rorschachs were frequent among people who became heavy smokers but it is also true that, with heavy smoking held constant, the presence of a coartated Rorschach seems related to inability to cut down or stop. Some of our scientists have smoked; they could easily stop. (They were not heavy smokers, of course.) Perhaps the scientists show that it is easier to quit when one's compeers do not smoke: none of our five writers can quit. Among our heavier smokers, the psychiatric labels of Sociable, Strong Basic Personality, and Practical characterize mostly people who can stop; Weak Basic Personality, Asocial, Lack of Purpose and Values, Introspective, Ideational, and Inhibited characterize mostly people who cannot. These are all small trends, however, and mostly occur in small numbers of cases. Yet these relationships seem worth mention because of their congruence with general theory. The variables related to smoking compulsively are mostly "need" variables, as the Bales theory would require.

A signal fact is this: ability to stop smoking is directly proportional to the number of months our subjects were fed from their

mothers' breast! The means march (.05) as follows: light smokers who can stop were weaned at 8.0 months; heavy smokers who can stop were weaned at 6.8 months; smokers, mostly heavy, who don't try to stop were weaned at 5.0 months; smokers, mostly heavy, who try to stop but cannot were weaned at 4.7 months. We had previously explored for a possible relationship between smoking and weaning or amount smoked and weaning, but it was not until we explored ability to stop smoking that the relationship to breastfeeding became so cleancut.

Many will wish to explain such a finding away. Certainly we would not argue that weaning is the cause of smoking. The argument must proceed through some *tertium quid*. In this we would agree with Linton (12), who stresses that such a crude datum as date of weaning or any other data "focussed primarily on actual technical operations, without a correspondingly detailed study of the maternal attitudes which accompanied these performances," is naive. The events of infant training are part of a broader pattern and symptomatic of it. "We must remember that these influences operate on the child from birth, and continue to operate on him for a long period of time . . . ." The congruence of infant training, childhood rearing, and adult values is shown by McClelland (15) for two contemporary American subcultures. So, in our data, late weaning was associated with those personality traits that are also related to ability to stop smoking. We do not have to postulate that infantile frustration and adult cigarette smoking are unmediated cause and effect.

Yet, it is a commonplace that people who stop smoking seek other oral gratifications. All the "small vices" we have correlated with smoking are oral vices: alcohol, coffee, and tobacco are taken in through the mouth. So is sugar, which Heath shows to be related to smoking. More directly to the point: Brozek (4) has experimentally documented the notion that men who give up smoking gain weight.

Theoretical reasons suggest that smoking should be correlated to psychoanalytic "orality." Our empirical data suggests that it is. Groups with the ability to stop smoking contain a smaller percentage of bottle babies. Thumbsucking was more commonly reported (close to .05) for men who continued to smoke. If one does not think of a psychoanalytic explanation, then a drive-reduction theory like that of Levy (11) seems to fit the data. Nor would these ideas be inconsistent with the Bales model: what they suggest is that these "deeper" needs have little or no effect on whether one smokes but great effect on how tenacious the habit, once adopted, may become.

In summary, the ability to stop smoking is grossly related to the amount of tobacco one has consumed. Good mental health, such that one has control over one's habits in general, seems to be relevant. So, apparently, is oral gratification received as an infant. This "orality" factor may be mediated in many ways, but its meaningfulness should not be overlooked.

## THE RELIABILITY OF THE FINDINGS

To check the reliability of some of these findings, we drew five per cent random samples of the class of 1958 and the class of 1961 from the files of the University Health Services. (*N*s were 55 and 58.) Every freshman fills in a medical questionnaire that contains an item about smoking as well as several items relevant to the variables related to smoking in the Study of Adult Development. Clearly, only the criterion of smoking in college (or near the start of college) was available. However, insofar as these checks bore out our findings, they showed them to be reproducible over a period (1939 to 1957) of almost twenty years.

Both samples showed an excess (.01) of public-school nonsmokers. In the first sample, the only one for which income data was available, it was noted again that income was not as good a predictor of smoking behavior as was type of preparatory school.

The relationship between seriousness of purpose and nonsmoking may be adumbrated by one datum. Within the public-school group, there is in both samples an excess of nonsmokers among those boys who characterize as Definite their pre-college choice of career. Similarly, in the sample for which income data is available, there is a strong tendency for the boys who are both Protestant and poor not to smoke. These would seem to be likely to be the lads for whom Harvard meant hard work and mobility. Or, viewed another way, these would

seem likely to include most of our members of Fundamentalist denominations.

In both freshmen samples, those boys who announce their intention of going into engineering or science tend not to smoke.

The examing physician rated these freshmen on "Personality Integration." For both samples, the percentage of smokers marches upwards as the ratings become less favorable. For the combined samples, the percentages of smokers are: for men rated "A," 0 per cent; for men rated "B," 23 per cent; for men rated "C," 37 per cent; and for men rated "D," 50 per cent. There were few A's and few D's. The contingency table made by combining A and B ratings in one column and combining C and D ratings in the other and setting these ratings against smoking and nonsmoking gives a chi square of 3.6, so $p$ is not quite as low as .05.

A similar rating by the examining physician was a prediction of College Adjustment. The percentages again march (though most clearly for the combined samples) as follows: among those rated "A," there were 13 per cent smokers; among the "Bs," 23 per cent; and among both "Cs" and "Ds," there were 50 per cent. Combining A and B and C and D, one gets a contingency table from which chi square is 8.0 and $p$ is less than .01. Both these findings from the physicians' ratings would seem to parallel the study findings that smoking went with psychiatric ratings suggesting poorer mental health.

The study finding that smoking and drinking are correlated was amply confirmed for the class of '58 but not so strongly shown in '61. In both samples, the tendency for drinking to precede smoking was noted.

On the whole, then, where we could cross-check the study findings, they seemed to hold up well. In this connection, it may be worth pointing out that a few of the patterns reported from the Harvard data have been found elsewhere. An unpublished survey of Yale seniors shows the public-private differences in smoking habits still pronounced at the end of the college career. The priority of drinking to smoking is noticeable. (This was, of course, first noted at Yale by Straus and Bacon.) The introjection of family standards shows very clearly as a source of smoking mores. We have already cited notable corroboration from market research done in England, where the association of nonsmoking with white-collar, middle-class identity was amply confirmed, as well as the definition of smoking as one of the minor vices. The same pattern is suggested by a national sample questioned about their smoking habits by the U. S. Bureau of Census (5). Results already cited from the University of Minnesota are consistent with our findings.

The one area in which no cross-validation has been possible has been that of psychodynamics, especially the findings from the Timed Rorschach. Since these personality patterns are very relevant to a theory of smoking, it is to be hoped that something may be done to check them.

## SUMMARY

A large part of what we have learned about the correlates of smoking habits may be made to fit a conceptual model like that of Bales. The fact that a man smokes or does not seems to be determined by whether or no he has been oriented to the habit as a result of his social milieu. Whether he becomes a heavy smoker or is unable to stop smoking seems determined by the usefulness of the smoking habit to his personal needs.

Nonsmokers tend to be lower-middle class in origin, upwardly mobile, earnest young men, bred in a work morality that is conducive to Inner Direction. Their parents and they themselves are often pious. They may pursue scientific or technical careers in many instances. Smokers are (in our data) likely to come from more privileged backgrounds, often entering business or humanistic careers, often having been raised in a Being or Being-in-Becoming orientation. Both subculture and the family as the mediator of the subculture are important determinants of whether and when the young man is oriented to the smoking habit.

Whether a smoker becomes a heavy smoker seems to depend on whether the habit serves many of his important needs. Very anxious or agitated men may adopt smoking as a tension reducer, but this use of smoking seems to be common only when the habit is already well established by other circumstances. Emotionally-constricted individuals seem to "take to" smoking with special eagerness.

Whether a man can alter his smoking habits seems to be most "deeply" determined. Efforts to quit or cut down seem to be normal,

since smoking is quite widely seen as a "small vice," even by smokers. Whether a man can succeed in these efforts is first of all a function of how much tobacco he has consumed in his lifetime. Certain personality variables are also relevant. In general, it is the needs that lead to heavy smoking after the habit has begun that also militate against quitting or cutting down. The social variables that were related to starting to smoke play little role in inability to quit. The effects of oral gratification at the breast may be relevant to control excessive smoking, once a great deal of tobacco has been consumed.

Where data were available, the study findings were cross-validated against samples from two Harvard classes, with generally encouraging results. Some confirmation of other findings has been had from unpublished studies at distant points and from a nationwide study by the Census Bureau.

Not all that we know about smoking fits into the proferred conceptual scheme. However, the scheme outlined makes our major findings hang together and may serve as a first hypothesis in further exploration of the psychology of smoking habits.

## REFERENCES


1. ALLPORT, G. The religion of the post-war college student. *J. Psychol.*, 1948, **25**, 3–33.
2. BALES, R. F. Social therapy for a social disorder—compulsive drinking. *J. soc. Issues*, 1945, **23**, 1–9.
3. BALES, R. F. Cultural differences in rates of alcoholism. *Quart. J. Stud. on Alcohol*, 1946, **6**, 480–499.
4. BROZEK, J., & KEYS, A. Changes of body weight in normal men who stop smoking cigarettes. *Science*, 1957, **125**, 1203.
5. HAENSZEL, W., SHIMKIN, M., & MILLER, H. Tobacco smoking patterns in the United States. Washington: U. S. Govt. Printing Office, 1956.
6. HEATH, C. W. *What people are*. Cambridge, Mass.: Harvard Univer. Press, 1945, p. 141.
7. HOBSEN, J., & HENRY, H. (Eds.). Patterns of smoking habits. *Hulton research studies of the British social pattern*. London: Hulton Press, 1948.
8. HOLLINGSHEAD, A. *Elmtown's youth*. New York: Wiley, 1949.
9. KLOPFER, B., & KELLY, D. *The Rorschach technique*. Yonkers-on-Hudson: World Book, 1942.
10. KLUCKHOHN, FLORENCE. Dominant and substitutive profiles of cultural orientations: their significance for the analysis of social stratification. *Soc. Forces*, 1950, **28**, 376–393.
11. LEVY, D. Thumb or finger sucking from the psychiatric angle. *Child Develpm.*, 1937, **8**, 99–101.
12. LINTON, R. *Culture and mental disorders*. Springfield, Ill.: Charles C Thomas, 1956.
13. MCARTHUR, C. Personalities of public and private school boys. *Harvard educ. Rev.*, 1954, **24** (4), 256–262.
14. MCARTHUR, C. Personality differences between middle and upper classes. *J. abnorm. soc. Psychol.*, 1955, **50** (2), 247–254.
15. MCCLELLAND, D., RINDLISBACHER, A., & DE CHARMS, R. Religious and other sources of parental attitudes toward independence training. In D. McClelland (Ed.), *Studies in motivation*. N. Y.: Appleton-Century-Crofts, 1955. Pp. 389–390.
16. MEAD, MARGARET. *Male and female*. New York: Mentor Books, 1955.
17. MILLER, D. C., & FORM, W. H. *Industrial sociology*. New York: Harper, 1951.
18. MILLS, C. W. *White collar*. New York: Oxford Univer. Press, 1951.
19. NIEBUHR, H. R. *The social sources of denominationalism*. New York: Meridian Books, 1957.
20. RIESMAN, D. *The lonely crowd*. New Haven: Yale Univer. Press, 1950.
21. RORSCHACH, H. *Psychodiagnostics*. Bern: Hans Huber, 1942.
22. STRAUS, R., & BACON, S. *Drinking in college*. New Haven: Yale Univer. Press, 1953.
23. WEBER, M. *The Protestant ethic*. (Trans. by Talcott Parsons) New York: Scribners, 1930.
24. WELLS, F. Mental factors in adjustment to higher education. *J. consult. Psychol.*, 1945, **9**, 51–56.
25. WELLS, F. Clinical psychometrics in higher education. *J. nat. Ass. Deans Women*, 1946, **9**, 51–56.
26. WELLS, F., & WOODS, W. Outstanding traits: In a selected college group, with some reference to career interests and war records. *Genet. Psychol. Monogr.*, 1946, **33**, 127–249.


Received September 27, 1957. Prior publication.

# CRITIQUE AND NOTES

## ATTITUDE CHANGE AND JUSTIFICATION FOR COMPLIANCE[1]

ARTHUR R. COHEN, JACK W. BREHM, AND WILLIAM H. FLEMING

*Yale University*

It has been demonstrated in a number of experiments (1, 3, 5, 6) that forcing a person to take an opinion position discrepant from what he privately believes results in change of private opinion toward the enforced one. The question then arises as to what variables affect the extent of private change under such conditions of compliance. This problem has recently been discussed by Festinger (2). According to his analysis, expressing an opinion discrepant from one's privately held position creates "dissonance" or psychological tension having drive character. The tension may, in this case, be reduced by changing one's private position so that it more nearly coincides with the position expressed.

While the direction of the private attitude change may thus be understood, what, according to dissonance theory, controls the magnitude of change? Both the number and importance of reasons in favor of and against expressing the discrepant opinion control the magnitude of dissonance. Given the expression of the discrepant opinion and holding constant the number of reasons for not expressing it, the greater the number and/or importance of reasons for expressing it, the less will be the dissonance. From these considerations we may state the major hypothesis: *with the discrepant expression and reasons for not making it held constant, the greater is the justification for making it, the less will be the magnitude of dissonance and consequent attitude change.*

In addition, with the reasons for and against taking the discrepant position held constant, the stronger the expression of the discrepant opinion, the greater will be the dissonance created and therefore, the greater will be the resultant attitude change.

An experiment by Kelman (4) has examined the effect on private attitudes of using specific nonsocial rewards to force compliance. The present study emphasizes another type of motivational induction, that of giving reasons unrelated to the attitude issue for expressing the discrepant opinion. In the present experiment, then, justification refers not to the attitude issue but to reasons for taking a view discrepant from one's private opinion.

[1] This report was prepared as part of the Yale Communications Research Program.

### METHOD

Subjects (*S*s) were 92 Yale College undergraduates who were approached in their classrooms and given an anonymous questionnaire. The questionnaire asked for their opinions on a salient issue, instructed them to write an essay *against* their position, obtained their opinions on the issue again, and measured their perceptions of the experimental situation.

The experiment was conducted shortly after a rumor that Yale would soon become coeducational had created considerable stir on the campus. Thus the issue of attitude toward coeducation at Yale was chosen as the salient attitude dimension. It was expected that most students would show an initial bias against coeducation.

The *S*s were first asked to indicate their attitudes toward admitting women undergraduates to Yale. The scale used was a "pick-two" device; from six choice points a ten-point scale was generated. *S*s were then asked to write an essay taking a stand *opposite* to the one initially indicated. This procedure was assumed to constitute the expression of an opinion position discrepant from the *S*s' private opinion. The instructions were the same for both experimental conditions except that the high justification group was given a number of reasons why the *S* should comply with the request to write the essay against his position.

In the *high justification* condition, the *S*s ($N=29$) were told the following: *E* appreciated their help since the study represented considerable research effort on his part; it was a seminar research paper which *E* had to get done; *E* had received a great deal of cooperation, interest, and willingness to help from other students; the ability to stand back and consider the other side of a complex issue was a real asset; considering the other side of the issue was only fair; taking the other side facilitated a good grasp of a complex issue; and the results of such a survey could be of interest to social scientists and the university authorities. They were given these cumulated arguments and urged to do the best possible job. *S*s in the *low justification* condition ($N=63$) were simply given the instructions common to both experimental groups, which asked them to write the most productive, creative, and imaginative essays possible, unequivocally against their own position. In order to minimize the pressure exerted to comply because of the very nature of the classroom situations, the *S*s were assured that they need not comply at all if they did not want to. Such perceived pressure due to classroom constraints might be expected to reduce dissonance and prevent variations in attitude change from emerging.

After writing their essays, the *S*s were again asked to indicate their attitude toward the issue in the light of additional thought, and also to respond to a post-

experimental questionnaire which attempted to measure the success of the experimental inductions.

## RESULTS

### *Check on the Experimental Inductions*

Data from the postexperimental questionnaire indicated that the experimental inductions were effective in creating differences between the two groups in justification for compliance. The low justification group perceived that less pressure to comply was exerted upon them than did the highs ($p < .02$), they were more self-motivated to write the essay than the highs ($p < .02$), and they showed fewer interfering responses than the highs ($p < .01$).[2]

### *Amount of Compliance and Attitude Change*

All *Ss* in both conditions complied. It may thus be inferred that the general force to comply was very compelling, and little attitude change is therefore to be expected. The average change in the high justification condition is .17 and that in the low is .46; both values are in the expected direction but neither is significantly different from no change.

### *Amount of Justification and Attitude Change*

The major hypothesis concerns the relationship between dissonance created by compliance and attitude change toward the assumed position. It was expected that the condition of low justification would create more dissonance and consequently more attitude change than would that of high justification. The obtained difference between the two conditions in mean change, .29, is in the predicted direction but is not statistically significant.

When the mean "before" opinions on the issue were examined for the high and low justification groups, the low group was found to be significantly more negative ($M = 6.50$ for the highs, 7.75 for the lows, $p < .02$). It was expected that negative attitudes toward coeducation would be shown more by upperclassmen who had internalized the Yale values than by lowerclassmen who had more recently come from secondary schools. In line with this expectation, the low justification group was found to contain far more upperclassmen than the high justification group, which contained relatively more lowerclassmen. The $p$ value is beyond the .001 level.

The implication of these relationships between experimental condition, extremity of attitude, and class at Yale becomes evident when the data on extremity of position and attitude change are examined. The results show that extremity of position on the attitude scale is inversely related to change in attitude ($p < .001$). Since more of the extremes are in the low justification group, probably due to the unequal distribution of upperclassmen in the two groups, and since extremity of position inhibits attitude change, it is difficult to test the hypothesis put forth. Therefore, it seems reasonable to eliminate the extremes (all those *Ss* who began at 10 plus the few who began at 1) from the sample in order to correct for this bias; the hypothesis thus is tested within a selected population ($N = 28$ for the low justification condition, $N = 17$ for the high justification condition). After selection, the means on the "before" opinion are now virtually identical: 5.80 for the highs, 5.72 for the lows.

One indication that the selection procedure is justified concerns the quality of the essays written. When the selected population is compared with the extremes on the number of original arguments given in their essays, 49 per cent of the selected population give one or more original arguments, while only 28 per cent of the extremes give such arguments ($p < .05$). The extremes thus appear to have been less involved in complying than the selected *Ss*; eliminating such differential involvement provides a better test of the hypothesis.

Comparison of the two selected groups on the effectiveness of the experimental inductions shows that they too differ on all three measures: the lows perceived less pressure, were more self-motivated, and had fewer interfering responses (all $p$ values beyond the .04 level).

When the major hypothesis is tested with the selected population, the lows show a mean change of .83, the highs a mean change of .22. The difference in amount of change is significant by one-tailed $t$ test at the .07 level. A tendency is therefore suggested by the data for greater change to occur in the condition where greater dissonance was postulated.

### *Strength of Induced Behavior*

Another assumed determinant of dissonance concerns the behavior which produces dissonance. It should hold that the better the quality of the essays written by the *Ss* against their original position, the more dissonance is created and, therefore, the greater should be the attitude change in line with the induced position. The essays were coded for number of original arguments, on the assumption that the improvisation of original arguments is a clear indication of the degree to which the *S* was actually involved in expressing the discrepant opinion. As such, the measure of original arguments represents the strength with which the discrepant opinion was

[2] Except where specifically mentioned, all statistical tests in this report are two-tailed tests.

expressed (3, 4). Among the selected *Ss*, a relationship between number of original arguments and attitude change was found only for the high justification group, the group where it was assumed little dissonance would be created. This relationship is significant at beyond the .001 level by chi-square test. For the low justification group where high dissonance and high attitude change occur, no additional dissonance is created by high quality of essay and, therefore, there are no variations in attitude change as a function of number of original arguments. In effect, where dissonance is not created by the experimental variables, it can occur as a function of individual variation in the strength of the induced behavior.

It might be argued that those who write high quality essays do so because they feel greater justification. This hypothesis is counter to our present reasoning since in the high justification condition, quality of essay is related to attitude change. However, no reliable relationships were found between quality of essay and perceived pressure (experimental checks on justification manipulations). Furthermore, it is clear that this hypothesis will not explain the major finding, namely, that attitude change tended to be greater where manipulated (and perceived) justification was least.

SUMMARY

This experiment has tentatively illustrated some of the processes involved in creating and reducing psychological dissonance and the implications of such dissonance for attitude change in compliance situations. It was hypothesized that increased justification for taking an opinion position discrepant from one's private opinion would lead to decreased dissonance and therefore decreased attitude change toward the discrepant position. *Ss* were asked to write an essay taking a stand opposite to their initial opinion on a given issue. One group of *Ss* was then given a number of reasons for compliance with the request (high justification condition). For another group (low justification condition) *Ss* were given no detailed justification for writing the essay. Attitude change was determined from a postexperimental questionnaire.

Three sources of dissonance were explored: (*a*) for the whole population, low dissonance and thus little attitude change were obtained because of the forceful nature of the experimental situation which brought general compliance; (*b*) high dissonance and therefore high attitude change tended to be associated with the low justification condition; and (*c*) in the high justification condition, variations in the strength of the induced behavior brought variations in dissonance and thus subsequent variations in attitude change toward the assumed position. These results should be accepted with caution since they arise from *post hoc* analyses.

REFERENCES

1. BURDICK, H. The compliant behavior of deviates under conditions of threat. Unpublished doctoral dissertation, Univer. of Minnesota, 1955.
2. FESTINGER, L. *Theory of cognitive dissonance.* Evanston: Row, Peterson, 1957.
3. JANIS, I. L., & KING, B. T. The influence of role playing in opinion change. *J. abnorm. and soc. Psychol.*, 1954, **49**, 211–218.
4. KELMAN, H. C. Attitude change as a function of response restriction. *Hum. Relat.*, 1953, **6**, 185–214.
5. MCBRIDE, D. The effects of public and private changes of opinion on intra-group communication. Unpublished doctoral dissertation, Univer. of Minnesota, 1955.
6. RAVEN, B. H. The effect of group pressures on opinion, perception and communication. Unpublished doctoral dissertation, Univer. of Michigan, 1953.

Received March 27, 1957.

# ANALITY, POLITICAL AGGRESSION, ACQUIESCENCE, AND QUESTIONNAIRE CONSTRUCTION

MAURICE L. FARBER

*University of Connecticut*

IN a recent paper (3) Rabinowitz attempts to demonstrate that Farber's (2) reported correlation between anality and political aggression can be explained away as a result of an acquiescence response set. There appear, however, to be serious defects in Rabinowitz's methods that render his conclusions doubtful. Christie, Havel, and Seidenberg's (1) recent cogent and detailed analysis of acquiescence is directly relevant; the present note will therefore be limited to brief comment on several features specific to Rabinowitz's study.

1. Rabinowitz's critique is based upon the construction of new Anality and Political Aggression scales which are scored negatively for these characteristics. In this difficult task (*vide* Christie *et al.*) he does not seem to have succeeded. Two items of

his five-item Anality scale (the second and fifth) and three of his four-item Political Aggression scale (first, second and fourth) appear highly doubtful. For example, is a "no" answer to "When I go to a restaurant I like to order foods which I have never tasted before" an anal reply? Or, is disagreement with the statement "The foreign policy of the United States should be based on the assumption that most other countries are our friends, not our enemies" an indication of political aggression? The rejection of a simplistic, soft-headed liberalism is not identical with active political hostility. Farber's anality items hew very closely to the Freudian formulations, and his political aggression items involve affirmation of actively aggressive feelings.

It is therefore not surprising that the correlation between Rabinowitz's scales and Farber's is but +.12 for anality and +.16 for political aggression. While no ultimate measure of validity is available for either set of scales, it is clear that the identically labeled scales are measuring something quite different. Since Rabinowitz's method requires psychologically equivalent scales that are worded in opposite directions, his failure to achieve these renders his critique irrelevant.

2. No item analysis is presented of Rabinowitz's scales, nor, perhaps more important, is there any indication that the discriminatory power of his items was tested. In view of the lack of theoretical relevance of some of his items as well as their possibly low discriminatory power, these omissions are serious ones.

3. Rabinowitz chose to drop Farber's modified Likert-type of scoring of the Political Aggression scale and to substitute simple agree-disagree categories. By so doing he precluded the possibility of isolating and properly weighting the genuinely aggressive response expressed in the "agree strongly" category, thereby dissipating much of the value of the scale. There is evidence, for example, that acquiescence responses are most frequent among those who agree slightly but relatively rare among those who agree strongly (1).

When one considers these methodological deficiencies in conjunction with Christie's demonstration of the small role played by acquiescence in items similar to those employed here, the lack of validity of Rabinowitz's strictures upon Farber's reported relationship is apparent.

REFERENCES

1. CHRISTIE, R., HAVEL, J., & SEIDENBERG, B. Is the F Scale irreversible? *J. abnorm. soc. Psychol*, 1958, **56**, 143–158.
2. FARBER, M. L. The anal character and political aggression. *J. abnorm. soc. Psychol.*, 1955, **51**, 486–489.
3. RABINOWITZ, W. Anality, aggression and acquiescence. *J. abnorm. soc. Psychol.*, 1957, **54**, 140–142.

Received April 10, 1957.

# INDEPENDENCE AND RESISTANCE TO PERCEPTUAL FIELD FORCES[1]

DOUGLAS N. JACKSON
*Pennsylvania State University*

THIS study was designed to test an hypothesis concerning consistencies in cognitive and personality organization. A second purpose was to provide further validation for previously observed relationships reflecting these hypothetical linkages. The general hypothesis may be stated as follows: individuals are consistent in their styles of response to field forces at different conceptual levels. Thus, *S*s who show personality characteristics consistent with independence from group pressure should also be able to resist forces in the perceptual field.

Jackson (4) has developed a reliable measure of the ability to resist reversals of perspective and has obtained a significant relationship between this measure and the tendency to disagree with F-scale items. Hypotheses provided by Witkin (7) and by Linton (6) regarding consistencies in perception and personality led the author to view this task as one in which *S* must cope actively with the perceptual field forces that cause figure and ground to reverse. Without necessarily invoking the hypothetical "brain fields" of the Gestaltist, it is possible to consider the inevitable and lawfully occurring reversals of perspective as manifestations of hypothetical forces in the visual field. Individual differences on this measure might be considered more generally as an indication of an *S*'s level of energy or activity in organizing perceptual and cognitive field forces, rather than submitting to

[1] This research was completed while the author was at The Menninger Foundation as a U. S. Public Health Service Clinical Research Fellow of the National Institute of Mental Health. The author wishes to thank Drs. N. H. Pronko, D. T. Herman, and G. E. Rice of the University of Wichita for making their students available as *S*s for this experiment. Dr. Riley Gardner's constructive comments concerning the manuscript are gratefully acknowledged. Portions of this paper were read at the Amer. Psychol. Ass., Chicago, 1956.

them. In testing this hypothesis, Jackson (5) found in a preliminary study that two criterion measures of this perceptual dimension correlated significantly and negatively with a measure of behavioral conformity and one of passivity on a group of males. The present study seeks to explore further the generality of this hypothetical dimension with a more economical group procedure.

## METHOD

To order people conveniently with respect to personality characteristics associated with independence of judgment, the personality items that empirically distinguished "independents" from "yielders" in conformity experiments by Barron (1) and by Crutchfield (2) were combined into an "inventory." Barron, using Asch's procedure, and Crutchfield, using a similar technique, placed *S* in a situation where experimentally created group pressure to conform was put in conflict with veridical perception and judgment. *S* had to choose between violating his own perception or publicly facing the prospect of being at variance with the judgment of the group. Behavior in these situations provides an index of an *S*'s susceptibility to external influence, of his capacity to withstand social field forces.

The inventory contained 45 true-false items. It was administered to groups of college students, 89 men and 54 women in all. The age range was 17–40, with a median of 20 years.

Two groups, "independents" and "yielders," were selected from the extremes of the distribution of "conformity inventory" scores. The *S*s were administered a group procedure adapted from one developed by Hochberg and McAlister (3) for determining the time one phase of a reversible figure appears to be in perspective. In this technique, a large India ink drawing of a Necker cube (11½ in. wide) was presented with explanation of the nature of the reversals of perspective. After a brief practice trial, *S*s were asked to attempt to hold in perspective the "up" phase (which was outlined) and to indicate by writing a "D" or a "U" which phase was in perspective just prior to the sounding of pure-tone, half-second "beeps." They were cautioned to keep their eyes on the cube throughout the experimental series. To assure an equivalent procedure for all groups, the instructions and the pure tone signals were recorded on tape and played through loudspeakers. There were two trials of two minutes each. Each trial contained 24 of the randomly spaced signals. The signals were randomized to prevent them from being synchronous with rhythmic reversals that occur in some *S*s. The score representing susceptibility to the visual field forces in this situation was the total number of "D" responses in the two two-minute trials.

## RESULTS

To test the hypothesis that "independents" would manifest significantly greater activity in coping with the visual field forces in the reversible figure, scores from the "conformity inventory" were tabulated and ranked. Because there was a small but significant difference between males and females on the inventory scores, the 27 per cent extreme groups were selected separately for each sex, thus, in effect, partialling out the influence of sex in making the major comparison. There was a total of 36 "independents" and 36 "yielders" selected from the extremes of the distribution of 143 *S*s. These extreme groups were then compared with respect to the perceptual scores, the number of "down" responses in the two trials of two minutes each. The differences are in the predicted direction; the "independents" showed a mean of 11.4 responses indicating that the "down" phase was in perspective (they were instructed to maintain the "up" phase) while the "yielders" had a mean of 17.5 "down" responses. A single classification analysis of variance between the two extreme groups indicated that this difference was significant ($F = 17.56$; $df = 70$; $p < .001$). Thus, evidence has been obtained for the hypothesis that "independents" may be more active in manipulating perceptual field forces.

Some interesting sidelights were obtained concerning some properties of the particular measures used in the study. For example, the corrected split-half reliability of the group measure of perception (.84) was found to compare fairly well with that obtained under conditions of individual administration, where it is usually above .90. The odd-even reliability of the conformity inventory, on the other hand, was quite low; after correction for length it was only .54 for the men and .44 for the women. Lack of homogeneity is often a serious problem in empirically derived indices such as the present one. It argues against the use of this scale for measuring attributes of particular individuals, although it may be of some use in discriminating broad groups for research purposes. On the inventory, females were found to show a slight but significant tendency ($p < .01$) to score in the "conforming" direction as compared with males. This finding is consistent with the type of developmental theory discussed by Witkin and his associates (7) and by others. Significant sex differences have not been encountered in the present measure of reversals of perspective, however.

It is interesting to note that 31 of the 45 conformity items must be answered *true* to be scored in the conforming direction. Acquiescent response-set may therefore be an important aspect of this measure. Such an interpretation is consistent conceptually with the independence-conformity dimension. Jackson's earlier finding (4) of a significant relationship between the perceptual measure and the F scale also may be consistent with this interpretation.

The evidence suggests that those who characteristically resist certain hypothetical perceptual field forces also have the personality character-

istics associated with independence of judgment, a form of resistance to social field forces. Those *Ss* who tended to answer personality items in the independent direction appeared to show greater activity in resisting reversals of perspective.

## REFERENCES


1. BARRON, F. Some personality correlates of independence of judgment. *J. Pers.*, 1953, **21**, 287–297.
2. CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191–198.
3. HOCHBERG, J., & MCALISTER, E. Quantitative approach to figural "goodness." *J. exp. Psychol.*, 1953, **46**, 361–364.
4. JACKSON, D. N. Cognitive energy level, acquiescence, and authoritarianism. *J. soc. Psychol.*, in press.
5. JACKSON, D. N. Stability in resistance to field forces. Unpublished doctoral dissertation, Purdue Univer., 1955.
6. LINTON, HARRIET B. Dependence on external influence: correlates in perception, attitudes, and judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 502–507.
7. WITKIN, H. A., LEWIS, HELEN B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, PEARL B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.


Received April 23, 1957.

# EXTRAPUNITIVENESS AND ETHNIC ATTITUDE

GERALD S. LESSER

*Hunter College*

ACKERMAN and Jahoda explain the complete absence of deep depression in their anti-Semitic psychoanalytic patients by stating, "The existence of an anti-Semitic reaction presupposes a tendency to blame the outside world rather than one's self, and dynamically, such a tendency is in contradiction to the overtly self-destructive trend of a genuine depressive" (1, p. 243). However, other studies (e.g., 3) have reported inconclusive results concerning the relationship between extrapunitiveness and prejudice. The present study examines the relationship between anti-Semitism and extrapunitiveness among both Jewish and non-Jewish boys.

## METHOD

The subjects were 20 white non-Jewish boys (ages 10-0 to 13-1) and 24 white Jewish boys (ages 10-2 to 13-2) drawn from one fifth-grade and two sixth-grade classes in two public schools in New Haven, Connecticut. The schools are in adjacent districts and the families constitute a relatively homogeneous upper lower-class group.

Extrapunitiveness was measured through the Children's Form of the Rosenzweig Picture-Frustration Study (5). The mean and standard deviation of the percentage of extrapunitive responses for the non-Jewish boys were 54.90 and 15.04 respectively, and for the Jewish boys were 55.24 and 13.62, respectively. These means are not significantly different.

Modified sociometric procedures have been frequently employed as a method for the indirect assessment of social and ethnic attitudes in children (e.g., 2, 4). Although variables other than prejudice also affect friendship selection, these sociometric techniques applied to the measurement of ethnic attitudes assume that the degree to which a child accepts or rejects the friendship of members of an ethnic group is related to the degree of prejudice he directs toward that group. In the present study, each child was asked to select his three best friends in his class and also to select the three boys that he wished were not in his class at all. An "anti-Semitism" score was obtained for each boy by subtracting the number of Jewish boys he chose for friends from the number of Jewish boys he rejected. A range of scores of +3 to −3 was obtained; the mean and standard deviation for non-Jewish boys were 0.00 and 1.42 respectively, and for the Jewish boys were −0.44 and 1.36, respectively. These means are not significantly different.

## RESULTS

Two Spearman rank-difference correlation coefficients (rho) were computed. For non-Jewish boys ($N = 20$), the correlation between extrapunitiveness and "anti-Semitism" is +.60 ($p < .01$). For Jewish boys ($N = 24$), the corresponding correlation is +.48 ($p < .01$). These coefficients are not significantly different.

## CONCLUSIONS

Previous studies (e.g., 3) have not found clear evidence of a relationship between extrapunitiveness and prejudice. The present study reports significant positive relationships between extrapunitiveness and anti-Semitism among both Jewish and non-Jewish boys.

Sarnoff (6) obtained a strong negative relationship between extrapunitiveness and anti-Semitism for a group of Jewish adults. For Jewish pre-adolescent boys, the present study found the reverse relationship, i.e., a significant positive relationship between extrapunitiveness and anti-Semitism.

REFERENCES

1. ACKERMAN, N. W., & JAHODA, MARIE. The dynamic basis of anti-Semitic attitudes. *Psychoanal. Quart.*, 1948, **17**, 240–260.
2. CRISWELL, JOAN H. A sociometric study of race cleavage in the classroom. *Arch. Psychol.*, 1939, No. 235.
3. LINDZEY, G. An experimental examination of the scapegoat theory of prejudice. *J. abnorm. soc. Psychol.*, 1950, **45**, 296–309.
4. MUSSEN, P. H. The reliability and validity of the Horowitz Faces Test. *J. abnorm. soc. Psychol.*, 1950, **45**, 504–506.
5. ROSENZWEIG, S., FLEMING, EDITH E., & ROSENZWEIG, LOUISE. The children's form of the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1948, **26**, 141–191.
6. SARNOFF, I. Identification with the aggressor: some personality correlates of anti-Semitism among Jews. *J. Pers.*, 1951, **20**, 199–218.

Received June 16, 1957.

# ANTICIPATED AUDIENCES AS DETERMINANTS OF RECALL

WILBUR SCHRAMM
*Stanford University*
AND WAYNE DANIELSON
*University of Wisconsin*

THIS is a study of the way in which reference groups affect how human beings handle the communications they receive. More precisely, it is a study of a *particular* reference group—the audience we expect to have to talk to concerning a given topic—and how our image of this group affects our handling of messages related to the specific topic.

Fifty-five years ago Cooley (1, especially pp. 58–62) developed the concept of "imaginary conversations." Much of our mental life, he said, is lived as a rehearsal of imaginary and anticipated conversations with persons who are important to us—or rather, with our images of those persons. "The mind," he said, "is not a hermit's cell, but a place of hospitality and intercourse. We have no higher life that is really apart from other people. It is by imagining them our personality is built up."

Many scholars have speculated as to just what effect this kind of mind life must have on perception, learning, and attitudes, but little or no experimenting had been done on it until in 1951. Grace (2) published a paper based on current developments in the theory of reference group and social perception. He showed subjects a large array of objects which ranged from intimate feminine to intimate masculine apparel. The experimental subjects were told to remember all these articles they could because they would later have to recall and repeat what they remembered. Some of them were told they would have to repeat them to a woman; others were not. *All* of them *did* report to a woman. Therefore the amount and nature of their recall could not be attributed to differential social constraint at naming intimate articles to a woman. There was no significant difference between the total number of items recalled by the two groups of subjects. However, those subjects who had been told they were to report to a woman tended to recall "feminine" articles significantly earlier in their lists. The experimenter's supposition was that these female-related items had been made more salient by the knowledge that the anticipated audience was to be female.

In 1956, Zimmerman and Bauer (3) reported an ingenious experiment which found clearer effects of an anticipated audience. Miss Zimmerman appeared before several classes, representing herself variously as speaking for an organization interested in improving teachers' salaries and as speaking for an organization concerned with saving taxpayers' money. She said her organization was seeking speakers to address its members on the subject of teachers' pay, and was also interested in how well people could remember arguments on the subject. She read a short passage to the classes, in half the classes favoring and in the other half opposing increases in teachers' salaries. She then asked the students to reproduce this passage, in writing, as accurately as possible. One week later she returned and asked them to write the passage again, as closely as they could remember it. Results were that at the end of a week, subjects remembered significantly more of the passage when they had been given a selection that was congruent with the attitude of the anticipated audience. That is, if a subject had been given a passage that opposed teachers' pay raises and was told that he was to write a speech for the Economy League, he remembered more of the passage than another subject whose passage *favored* pay raises and who was also to speak to

the Economy League. The experimenters also found that subjects especially concerned with the audience (e.g., journalism students) showed more effect of this anticipated audience than subjects chiefly concerned with the topic (e.g., education students).

The present experimenters were planning a study of this problem at the same time as Miss Zimmerman was collecting her data. When they found out about her work they modified their design somewhat in order to replicate the essential feature of the Zimmerman-Bauer experiment and also to include an aspect of the relationship that was not included in the earlier work: the effect of an anticipated audience on attitude. The latter aspect of this study is not reported here because the results (which did not confirm any simple relationship of attitude change to anticipated audience) call for further research before they can be interpreted confidently.

## METHOD

This experiment, then, was based on the supposition that a large percentage of the communications we receive during life are such as to call up an anticipated audience—that is, some reference group or individual to whom we think we may have to retransmit the message, or with whom we may want to talk it over, or with whom we may conduct an imaginary conversation concerning it. Given that situation, it was hypothesized that a person tends to remember the parts of a received communication which he believes will appeal to the interests or be consonant with the relevant attitudes of the anticipated audience.

In order to test this hypothesis, approximately 200 Stanford speech students were divided into three groups. A control group received attitude tests as below, but no experimental communications. The other two groups were given a two-sided set of arguments for and against lowering the voting age to 18. These arguments were equal in number for each side and were matched by judges. These test groups were told that they were to prepare speeches in competition for a prize. One test group was told that the prize was to be given by "The Society for Constitutional Government —dedicated to preserving the U. S. Constitution unchanged" and hence opposed to any change in voting age; the other that the prize was to be given by "The Association for Lowering the Voting Age to 18—If they can fight, they can vote." To reinforce this setting, the name of the mythical organization for which the student was expected to write and speak appeared on the page with the arguments, although the same set of arguments was used under both names.

TABLE 1

RECALL OF ARGUMENTS IN RELATION TO AN ANTICIPATED AUDIENCE

| | Number of Arguments Recalled *For* (Mean) | Number of Arguments Recalled *Against* (Mean) | Mean Difference | Variance of Difference |
|---|---|---|---|---|
| Audience *For* Lowering Voting Age ($N = 76$) | 3.57 | 3.12 | 0.45 | 1.83 |
| Audience *Against* Lowering Voting Age ($N = 80$) | 3.80 | 4.19 | −0.39 | 1.64 |

CR for net difference (DF − DA) equals 3.98; $p$ equals .0008 (2-tailed test).

An attitude test was administered first. Then the students were told about the prize contest, given the arguments to read, and told to remember them because they would be useful later when they would write the speech. One week later, the teacher administered a memory test on the arguments, and also repeated the attitude test. Later the students did write speeches on this subject, and small prizes were given for the best ones.

The groups were randomly assigned and found to be closely equated on the obviously relevant measures—age, year in college, sex, and major subject.

## RESULTS

It was found that the subjects remembered significantly more arguments on the side they were supposed to address, as Table 1 shows. Thus it would appear that the recall of arguments from a two-sided group of arguments is significantly greater after one week for those arguments which agree with the supposed position of an audience with which the subject expects to have to discuss the topic. The hypothesis, and the Zimmerman-Bauer conclusion (3) must be regarded as confirmed.

## REFERENCES

1. COOLEY, C. H. *Human nature and the social order.* New York: Scribner, 1902.
2. GRACE, H. A. Effects of different degrees of knowledge about an audience on the content of communication. *J. soc. Psychol.*, 1951, **34**, 111–124.
3. ZIMMERMAN, CLAIRE, & BAUER, R. A. The influence of an audience on what is remembered. *Publ. Opin. Quart.*, 1956, **20**, 238–248.

Received August 13, 1957. Prior publication.

# ERRATA

In the article, "The Relationship Between Advanced Age and Mental Abilities", by Charles R. Strother, K. Warner Schaie, and Paul Horst, which appeared in the September, 1957 issue of this Journal (*J. Abnorm. Soc. Psychology*, 1957, **55**, 166–170), the cuts for Figs. 1 and 3 were erroneously transposed in printing. The graph labelled Fig. 1 is derived from the data shown in Table 4 and represents the *corrected* T-score values. The graph labelled Fig. 3 is derived from the data in Table 1 and shows the *uncorrected* values.

---

In the article by E. H. Schein, W. F. Hill, H. L. Williams, and A. Lubin, "Distinguishing Characteristics of Collaborators and Resisters Among American Prisoners of War," this Journal, 1957, **55**, 197–201, the final sentence of the first main paragraph on p. 199 should read, "No comparisons between resisters and *collaborators* are shown, as none of these were significant."

# JOURNAL OF MENTAL DEFICIENCY RESEARCH

*Chairman*
L. S. Penrose

*Editor*
B. W. Richards
St. Lawrence's Hospital, Caterham, Surrey

*Publishing Manager*
A. Highfield
10, Shendon Way, Sevenoaks, Kent

A Journal with a medical bias, but accepting original contributions on all aspects of mental deficiency. Appears twice a year.

Single copy 7/6 or 75 cents.
Yearly 15/- or $1.50

Subscriptions to Publishing Manager
Contributions to Editor

## RONALD Books...

### Social Psychology—*2nd Edition*

**S. Stansfeld Sargent, Ph.D.; Robert C. Williamson, Ph. D.** *Just Published.* An introduction to the study of human relations. Offers a systematic, integrative approach to social behavior. Now completely up to date, book covers culture and personality, social attitudes and groups, communication, development of the self, etc. *33 illus., tables, 639 pp.*

### Intensive Group Psychotherapy

**George R. Bach, Ph.D.** Presents new conceptual and clinical tools beyond those developed in individual therapy. This practical book fully discusses group therapy and the dynamic forces influencing it. Details on: group size, time and space needs, criteria for selecting patients, etc. *Illus.; 208 pp.*

### Interpersonal Diagnosis of Personality

**Timothy Leary, Ph.D.** Pioneering book provides an original personality theory, a series of techniques for measuring interpersonal expression at different personality levels, and an empirical method for applying theory and technique clinically. *120 illus., tables; 518 pp.*

### Energy and Structure in Psychoanalysis

**Kenneth Mark Colby, M.D.** An original study of the basic theory of psychoanalysis, examining the postulates of energy and structure, and showing their usefulness and drawbacks in theoretical problems. *Illus.*

### The Sentence Completion Method

**Amanda R. Rohde.** Examines the sentence completion method as an important diagnostic and clinical tool. Describes the method, its rationale, administration, and scoring. Twenty-three documented, scored, and interpreted cases. *28 illus., tables, 301 pp.*

### Psychotherapy and Culture Conflict

**Georgene Seward, Ph.D.** A thorough analysis of the psychodynamics peculiar to the members of a variety of subcultures in the U. S. Includes four case studies contributed by **Judd Marmor, M.D.**, from his own practice. *255 pp.*

THE RONALD PRESS COMPANY
15 East 26th St., New York 10

## PERSONAL ADJUSTMENT: An Approach through the Study of Healthy Personality

*By* SIDNEY M. JOURARD, *Ph.D., Associate Professor of Psychology, University of Florida.* Written from a positive point of view, this unique work discusses the factors determining the healthy adjusted personality. The author outlines each aspect of personality and compares the normal, unhealthy and healthy versions of the trait in question. The sections on expositions of interpersonal relations, love, the self-structure, and conscience are new to the field.

A pre-publication reader comments: "I think it is indeed a new approach to mental hygiene and a far better one than what we now have available."

1958, 462 pages, $5.50

## INTRODUCTORY CLINICAL PSYCHOLOGY

*By* SOL L. GARFIELD, *Chief of the Clinical Psychology Division and Associate Professor in the Psychiatric Institute, University of Nebraska Medical School.*

1957, 469 pages, $6.00

## THE PSYCHOLOGY OF ADOLESCENCE

*By* ARTHUR T. JERSILD, *Professor of Education, Psychological Foundations and Services Department, Teachers College, Columbia University.*

1957, 438 pages, $5.00

## THE PSYCHOLOGY OF RELIGION: An Introduction to Religious Experience and Behavior

*By* WALTER HOUSTON CLARK, *Professor of Psychology and Dean of the Hartford School of Religious Education, Hartford Seminary Foundation.*

Ready Spring 1958

## DIFFERENTIAL PSYCHOLOGY: Individual and Group Differences in Behavior, Third Edition

*By* ANNE ANASTASI, *Professor of Psychology, Graduate School, Fordham University.*

Spring, 1958

*The Macmillan Company*

60 FIFTH AVENUE, NEW YORK 11, N. Y.